

University of Groningen

Lyapunov Criterion for Stochastic Systems and Its Applications in Distributed Computation

Qin, Yuzhen; Cao, Ming; Anderson, Brian D. O.

Published in:
IEEE-Transactions on Automatic Control

DOI:
[10.1109/TAC.2019.2910948](https://doi.org/10.1109/TAC.2019.2910948)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Qin, Y., Cao, M., & Anderson, B. D. O. (2020). Lyapunov Criterion for Stochastic Systems and Its Applications in Distributed Computation. *IEEE-Transactions on Automatic Control*, 65(2), 546-560. [8691522]. <https://doi.org/10.1109/TAC.2019.2910948>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Lyapunov Criterion for Stochastic Systems and Its Applications in Distributed Computation

Yuzhen Qin, *Student Member, IEEE*, Ming Cao, *Senior Member IEEE*,
and Brian D. O. Anderson, *Life Fellow, IEEE*

Abstract—This paper presents new sufficient conditions for convergence and asymptotic or exponential stability of a stochastic discrete-time system, under which the constructed Lyapunov function always decreases in expectation along the system's solutions after a finite number of steps, but without necessarily strict decrease at every step, in contrast to the classical stochastic Lyapunov theory. As the first application of this new Lyapunov criterion, we look at the product of any random sequence of stochastic matrices, including those with zero diagonal entries, and obtain sufficient conditions to ensure the product almost surely converges to a matrix with identical rows; we also show that the rate of convergence can be exponential under additional conditions. As the second application, we study a distributed network algorithm for solving linear algebraic equations. We relax existing conditions on the network structures, while still guaranteeing the equations are solved asymptotically.

I. INTRODUCTION

STABILITY analysis for stochastic dynamical systems has always been an active research field. Early works have shown that stochastic Lyapunov functions play an important role, and to use them for discrete-time systems, a standard procedure is to show that they decrease in *expectation* at every time step [1]–[4]. Properties of supermartingales and LaSalle's arguments are critical to establish the related proofs. However, most of the stochastic stability results are built upon a crucial assumption, which requires that the stochastic dynamical system under study is Markovian (see e.g., [1]–[3], [5]), and very few of them have reported bounds for the convergence speed.

More recently, with the fast development of network algorithms, more and more distributed computational processes are carried out in networks of coupled computational units. Such dynamical processes are usually modeled by stochastic discrete-time dynamical systems since they are usually under inevitable influences from random changes of network structures [6]–[9], communication delay and noise [10]–[12], and asynchronous updating events [13], [14]. So there is great need in further developing Lyapunov theory for stochastic

dynamical systems, in particular in the setting of network algorithms for distributed computation. And this is exactly the aim of this paper.

We aim at further developing the Lyapunov criterion for stochastic discrete-time systems. Motivated by the concept of *finite-step Lyapunov functions* for deterministic systems [15]–[17], we propose to define a *finite-step stochastic Lyapunov function*, which decreases in expectation, not necessarily at every step, but after a finite number of steps. The associated new Lyapunov criterion not only enlarges the range of choices of candidate Lyapunov functions but also implies that the systems that it can be used to analyze do not need to be Markovian. An additional advantage of using this new criterion is that we are enabled to construct conditions to guarantee exponential convergence and estimate convergence rates.

We then apply the finite-step stochastic Lyapunov function to study two distributed computation problems arising in some popular network algorithmic settings. In distributed optimization [18], [19] and other distributed coordination algorithms [7], [20]–[22], one frequently encounters the need to prove convergence of inhomogeneous Markov chains, or equivalently the convergence of backward products of random sequences of stochastic matrices $\{W(k)\}$. Most of the existing results assume exclusively that all the $W(k)$ in the sequence have all positive diagonal entries, see e.g., [23]–[25]. This assumption simplifies the analysis of convergence significantly; moreover, without this assumption, the existing results do not always hold. For example, from [7], [22] one knows that the product of $W(k)$ converges to a rank-one matrix almost surely if exactly one of the eigenvalues of the expectation of $W(k)$ has the modulus of one, which can be violated if $W(k)$ has zero diagonal elements. Note also that most of the existing results are confined to special random sequences, e.g., independently distributed sequences [22], stationary ergodic sequences [7], or independent sequences [26], [27]. Using the new Lyapunov criterion in this paper, we work on more general classes of random sequences of stochastic matrices without the assumption of non-zero diagonal entries. We show that if there exists a fixed length such that the product of any successive subsequence of matrices of this length has the *scrambling* property (a standard concept, but it will be defined subsequently) with positive probability, the convergence to a rank-one matrix for the infinite product can be guaranteed almost surely. We also prove that the convergence can be exponentially fast if this probability is lower bounded by some positive number, and the greater the lower bound is, the faster the convergence becomes. For some particular random

Y. Qin and M. Cao are with the Institute of Engineering and Technology, Faculty of Science and Engineering, University of Groningen, Groningen, the Netherlands (y.z.qin, m.cao@rug.nl). B.D.O. Anderson is with School of Automation, Hangzhou Dianzi University, Hangzhou, 310018, China, and Data61-CSIRO and Research School of Electrical, Energy and Materials Engineering, Australian National University, Canberra, ACT 2601, Australia (Brian.Anderson@anu.edu.au).

The work of Cao was supported in part by the European Research Council (ERC-CoG-771687) and the Netherlands Organization for Scientific Research (NWO-vidi-14134). The work of B.D.O. Anderson was supported by the Australian Research Council (ARC) under grants DP-130103610 and DP-160104500, and by Data61-CSIRO.

sequences, we further relax this “scrambling” condition. If the random sequence is driven by a *stationary* process, the almost sure convergence can be ensured as long as the product of any successive subsequence of finite length has positive probability to be indecomposable and aperiodic (SIA). The exponential convergence rate follows without other assumptions if the random process that governs the evolution of the sequence is a *stationary ergodic* process.

As the second application of the finite-step stochastic Lyapunov functions, we investigate a distributed algorithm for solving linear algebraic equations of the form $Ax = b$. The equations are solved in parallel by n agents, each of whom just knows a subset of the rows of the matrix $[A, b]$. Each agent recursively updates its estimate of the solution using the current estimates from its neighbors. Recently several solutions under different sufficient conditions have been proposed [28]–[30], and in particular in [30], the sequence of the neighbor relationship graphs $\mathcal{G}(k)$ is required to be repeated jointly strongly connected. We show that a much weaker condition is sufficient to solve the problem almost surely, namely the algorithm in [30] works if there exists a fixed length such that any subsequence of $\{\mathcal{G}(k)\}$ at this length is jointly strongly connected with positive probability.

The remainder of this paper is organized as follows. In Section II, we define the finite-step stochastic Lyapunov functions. Products of random sequences of stochastic matrices are studied in Section III; in Section IV we look into in particular the asynchronous implementation issues as an application of Section III. Finally, we study in Section V a distributed approach for solving linear equations. Brief concluding remarks appear in Section VI.

Notation: Throughout this paper, \mathbb{N}_0 denotes the sets of non-negative integers, \mathbb{N} the collection of positive integers, and \mathbb{R}^q the real q -dimensional vector space. Moreover, we let $\mathbf{1}$ be the vector consisting of all ones, and let $\mathbf{N} = \{1, 2, \dots, n\}$. Given a vector $x \in \mathbb{R}^n$, x^i denotes the i th element of x . Let $\|\cdot\|$, $p \geq 1$, be any p -norm. A continuous function $h(x) : [0, a) \rightarrow [0, \infty)$ is said to belong to class \mathcal{K} if it is strictly increasing and $h(0) = 0$. For any two events A, B , the conditional probability $\Pr[A|B]$ denotes the probability of A given B .

II. FINITE-STEP STOCHASTIC LYAPUNOV FUNCTIONS

Consider a stochastic discrete-time system described by

$$x_{k+1} = f(x_k, y_{k+1}), \quad k \in \mathbb{N}_0, \quad (1)$$

where $x_k \in \mathbb{R}^n$, and $\{y_k : k \in \mathbb{N}\}$ is a \mathbb{R}^d -valued stochastic process on a probability space $(\Omega, \mathcal{F}, \Pr)$. Here $\Omega = \{\omega\}$ is the sample space; \mathcal{F} is a set of events which is a σ -field; $\Pr : \mathcal{F} \rightarrow [0, 1]$ is a function that assigns probabilities to events; y_k is a measurable function mapping Ω into the state space $\Omega_0 \subseteq \mathbb{R}^d$, and for any $\omega \in \Omega$, $\{y_k(\omega) : k \in \mathbb{N}\}$ is a realization of the stochastic process $\{y_k\}$ at ω . Let $\mathcal{F}_k = \sigma(y_1, \dots, y_k)$ for $k \geq 1$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$, so that evidently $\{\mathcal{F}_k\}, k = 1, 2, \dots$, is an increasing sequence of σ -fields. Following [31], we consider a constant initial condition $x_0 \in \mathbb{R}^n$ with probability one. It then can be observed that the solution to (1), $\{x_k\}$, is a \mathbb{R}^n -valued stochastic process

adapted to \mathcal{F}_k . The randomness of y_k can be due to various reasons, e.g., stochastic disturbances or noise. Note that (1) becomes a stochastic switching system if $f(x, y) = g_y(x)$, where y maps Ω into the set $\Omega_0 := \{1, \dots, p\}$, and $\{g_p(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n, p \in \Omega_0\}$ is a given family of functions.

A point x^* is said to be an *equilibrium* of system (1) if $f(x^*, y) = x^*$ for any $y \in \Omega_0$. Without loss of generality, we assume that the origin $x = 0$ is an equilibrium. Researchers have been interested in studying the limiting behavior of the solution $\{x_k\}$, i.e., when and to where x_k converges as $k \rightarrow \infty$. Most noticeably, Kushner developed classic results on stochastic stability by employing stochastic Lyapunov functions [1]–[3]. We introduce some related definitions before recalling some Kushner’s results. Following [32, Sec. 1.5.6] and [33], we first define convergence and exponential convergence of a sequence of random variables.

Definition 1 (Convergence). *A random sequence $\{x_k \in \mathbb{R}^n\}$ in a sample space Ω converges to a random variable x almost surely if $\Pr[\omega \in \Omega : \lim_{k \rightarrow \infty} \|x_k(\omega) - x\| = 0] = 1$. The convergence is said to be exponentially fast with a rate no slower than γ^{-1} for some $\gamma > 1$ independent of ω if $\gamma^k \|x_k - x\|$ almost surely converges to y for some finite $y \geq 0$. Furthermore, let $\mathcal{D} \subset \mathbb{R}^n$ be a set; a random sequence $\{x_k\}$ is said to converge to \mathcal{D} almost surely if $\Pr[\omega \in \Omega : \lim_{k \rightarrow \infty} \text{dist}(x_k(\omega), \mathcal{D}) = 0] = 1$, where $\text{dist}(x, \mathcal{D}) := \inf_{y \in \mathcal{D}} \|x - y\|$.*

Here “almost surely” is exchangeable with “with probability one”, and we sometimes use the shorthand notation “a.s.”. We now introduce some stability concepts for stochastic discrete-time systems analogous to those in [5] and [34] for continuous-time systems¹.

Definition 2. *The origin of (1) is said to be:*

- 1) stable in probability if $\lim_{x_0 \rightarrow 0} \Pr[\sup_{k \in \mathbb{N}} \|x_k\| > \varepsilon] = 0$ for any $\varepsilon > 0$;
- 2) asymptotically stable in probability if it is stable in probability and moreover $\lim_{x_0 \rightarrow 0} \Pr[\lim_{k \rightarrow \infty} \|x_k\| = 0] = 1$;
- 3) exponentially stable in probability if for some $\gamma > 1$ independent of ω , $\lim_{x_0 \rightarrow 0} \Pr[\lim_{k \rightarrow \infty} \|\gamma^k x_k\| = 0] = 1$;

Definition 3. *For a set $\mathcal{Q} \subseteq \mathbb{R}^n$ containing the origin, the origin of (1) is said to be:*

- 1) locally a.s. asymptotically stable in \mathcal{Q} (globally a.s. asymptotically stable, respectively) if starting from $x_0 \in \mathcal{Q}$ ($x_0 \in \mathbb{R}^n$, respectively) all the sample paths x_k stay in \mathcal{Q} (\mathbb{R}^n , respectively) for all $k \geq 0$ and converge to the origin almost surely;
- 2) locally a.s. exponentially stable in \mathcal{Q} (globally a.s. exponentially stable, respectively) if it is locally (globally, respectively) a.s. asymptotically stable and the convergence is exponentially fast.

Now let us recall some Kushner’s results on convergence and stability, where stochastic Lyapunov functions have been used.

¹Note that 1) and 2) of Definition 2 follow from the definitions in [5, Chap. 5], in which an arbitrary initial time s rather than just 0 is actually considered; we define 3) following the same lines as 1) and 2). In Definition 3, 1) follows from the definitions in [34], and we define 2) following the same lines as 1).

Lemma 1 (Asymptotic Convergence and Stability). *For the stochastic discrete-time system (1), let $\{x_k\}$ be a Markov process. Let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous positive definite and radially unbounded function. Define the set $\mathcal{Q}_\lambda := \{x : 0 \leq V(x) < \lambda\}$ for some $\lambda > 0$, and assume that*

$$\mathbb{E}[V(x_{k+1}) | x_k] - V(x_k) \leq -\varphi(x_k), \forall k, \quad (2)$$

where $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous and satisfies $\varphi(x) \geq 0$ for any $x \in \mathcal{Q}_\lambda$. Then the following statements apply:

i) for any initial condition $x_0 \in \mathcal{Q}_\lambda$, x_k converges to $\mathcal{D}_1 := \{x \in \mathcal{Q}_\lambda : \varphi(x) = 0\}$ with probability at least $1 - V(x_0)/\lambda$ [3];

ii) if moreover $\varphi(x)$ is positive definite on \mathcal{Q}_λ , and $h_1(\|s\|) \leq V(s) \leq h_2(\|s\|)$ for two class \mathcal{K} functions h_1 and h_2 , then $x = 0$ is asymptotically stable in probability [35, Theorem 7.3].

Lemma 2 (Exponential Convergence and Stability). *For the stochastic discrete-time system (1), let $\{x_k\}$ be a Markov process. Let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous nonnegative function. Assume that*

$$\mathbb{E}[V(x_{k+1}) | x_k] - V(x_k) \leq -\alpha V(x_k), \quad 0 < \alpha < 1. \quad (3)$$

Then the following statements apply:

i) for any given x_0 , $V(x_k)$ almost surely converges to 0 exponentially fast with a rate no slower than $1 - \alpha$ [2, Th. 2, Chap. 8], [35];

ii) if moreover V satisfies $c_1\|x\|^a \leq V(x) \leq c_2\|x\|^a$ for some $c_1, c_2, a > 0$, then $x = 0$ is globally a.s. exponentially stable [35, Theorem 7.4].

To use these two lemmas to prove asymptotic (or exponential) stability for a stochastic system, the critical step is to find a stochastic Lyapunov function such that (2) (respectively, (3)) holds. However, it is not always obvious how to construct such a stochastic Lyapunov function. We use the following toy example to illustrate this point.

Example 1. Consider a randomly switching system described by $x_k = A_{y_k} x_{k-1}$, where y_k is the switching signal taking values in a finite set $\mathcal{P} := \{1, 2, 3\}$, and

$$A_1 = \begin{bmatrix} 0.2 & 0 \\ 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0.8 \end{bmatrix}, A_3 = \begin{bmatrix} 1 & 0 \\ 0 & 0.6 \end{bmatrix}.$$

The stochastic process $\{y_k\}$ is described by a Markov chain with initial distribution $v = \{v_1, v_2, v_3\}$. The transition probabilities are described by a transition matrix

$$\pi = \begin{bmatrix} 0 & 0.4 & 0.6 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

whose ij th element is defined by $\pi_{ij} = \Pr[y_{k+1} = j | y_k = i]$. Since $\{y_k\}$ is not independent and identically distributed, the process $\{x_k\}$ is not Markovian. Nevertheless, we might conjecture that the origin is globally a.s. exponentially stable. In order to try to prove this, we might choose a stochastic Lyapunov function candidate $V(x) = \|x\|_\infty$, but the existing results introduced in Lemma 2 cannot be used since $\{x_k\}$ is not Markovian. Moreover, by calculation we can only observe that $\mathbb{E}[V(x_{k+1}) | x_k, y_k] \leq V(x_k)$ for any y_k , which implies

that (3) is not necessarily satisfied. Thus $V(x)$ is not an appropriate stochastic Lyapunov function for which Lemma 2 can be applied. As it turns out however, the same $V(x)$ can be used as a Lyapunov function to establish exponentially stability via the alternative criterion set out subsequently. \triangle

It is difficult, if not impossible, to construct a stochastic Lyapunov function, especially when the state of the system is not Markovian. So it is of great interest to generalize the results in Lemmas 1 and 2 such that the range of choices of candidate Lyapunov functions can be enlarged. For deterministic systems, Aeyels et al. have introduced a new Lyapunov criterion to study asymptotic stability of continuous-time systems [15]; a similar criterion has also been obtained for discrete-time systems, and the Lyapunov functions satisfying this criterion are called *finite-step Lyapunov functions* [16], [17]. A common feature of these works is that the Lyapunov function is required to decrease along the system's solutions after a finite number of steps, but not necessarily at every step. We now use this idea to construct stochastic finite-step Lyapunov functions, a task which is much more challenging compared to the deterministic case due to the uncertainty present in stochastic systems. The tools for analysis are totally different from what are used for deterministic systems. We will exploit supermartingales and their convergence property, as well as the Borel-Cantelli Lemma; these concepts are introduced in the two following lemmas.

Lemma 3 ([36, Sec. 5.2.9]). *Let the sequence $\{X_k\}$ be a nonnegative supermartingale with respect to $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$, i.e., suppose: (i) $\mathbb{E}X_n < \infty$; (ii) $X_k \in \mathcal{F}_k$ for all k ; (iii) $\mathbb{E}(X_{k+1} | \mathcal{F}_k) \leq X_k$. Then there exists some random X such that $X_k \xrightarrow{a.s.} X, k \rightarrow \infty$, and $\mathbb{E}X \leq \mathbb{E}X_0$.*

Lemma 4 ([2, P.192]). *Let $\{X_k\}$ be a nonnegative random sequence. If $\sum_{k=0}^{\infty} \mathbb{E}X_k < \infty$, then $X_k \xrightarrow{a.s.} 0$.*

We are now ready to present our first main result on stochastic convergence and stability.

Theorem 1. *For the stochastic discrete-time system (1), let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous nonnegative and radially unbounded function. Define the set $\mathcal{Q}_\lambda := \{x : V(x) < \lambda\}$ for some $\lambda > 0$, and assume that*

a) $\mathbb{E}[V(x_{k+1}) | \mathcal{F}_k] - V(x_k) \leq 0$ for any k such that $x_k \in \mathcal{Q}_\lambda$;

b) there is an integer $T \geq 1$, independent of ω , such that for any k , $\mathbb{E}[V(x_{k+T}) | \mathcal{F}_k] - V(x_k) \leq -\varphi(x_k)$, where $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous and satisfies $\varphi(x) \geq 0$ for any $x \in \mathcal{Q}_\lambda$.

Then the following statements apply:

i) for any initial condition $x_0 \in \mathcal{Q}_\lambda$, x_k converges to $\mathcal{D}_1 := \{x \in \mathcal{Q}_\lambda : \varphi(x) = 0\}$ with probability at least $1 - V(x_0)/\lambda$;

ii) if moreover $\varphi(x)$ is positive definite on \mathcal{Q}_λ , and $h_1(\|s\|) \leq V(s) \leq h_2(\|s\|)$ for two class \mathcal{K} functions h_1 and h_2 , then $x = 0$ is asymptotically stable in probability.

Proof. Before proving i) and ii), we first show that starting from $x_0 \in \mathcal{Q}_\lambda$ the sample paths $x_k(\omega)$ stay in \mathcal{Q}_λ with probability at least $1 - V(x_0)/\lambda$ if Assumption a) is satisfied. This has been proven in [2, p. 196] by showing that

$$\Pr[\sup_{k \in \mathbb{N}} V(x_k) \geq \lambda] \leq V(x_0)/\lambda. \quad (4)$$

Let $\bar{\Omega}$ be a subset of the sample space Ω such that for any $\omega \in \bar{\Omega}$, $x_k(\omega) \in \mathcal{Q}_\lambda$ for all k . Let J be the smallest $k \in \mathbb{N}$ (if it exists) such that $V(x_k) \geq \lambda$. Note that, this integer J does not exist when $x_k(\omega)$ stays in \mathcal{Q}_λ for all k , i.e., when $\omega \in \bar{\Omega}$.

We first prove i) by showing that the sample paths staying in the \mathcal{Q}_λ converge to \mathcal{D}_1 with probability one, i.e., $\Pr[x_k \rightarrow \mathcal{D}_1 | \bar{\Omega}] = 1$. Towards this end, define a new function $\tilde{\varphi}(x)$ such that $\tilde{\varphi}(x) = \varphi(x)$ for $x \in \mathcal{Q}_\lambda$, and $\tilde{\varphi}(x) = 0$ for $x \notin \mathcal{Q}_\lambda$. Define another random process $\{\tilde{z}_k\}$. If J exists, when $J > T$ let

$$\tilde{z}_k = x_k, k < J - T, \quad \tilde{z}_k = \epsilon, k \geq J - T,$$

where ϵ satisfies $V(\epsilon) = \bar{\lambda} > \lambda$; when $J \leq T$, let $\tilde{z}_k = \epsilon$ for any $k \in \mathbb{N}_0$. If J does not exist, we let $\tilde{z}_k = x_k$ for all $k \in \mathbb{N}_0$. Then it is immediately clear that $\mathbb{E}[V(\tilde{z}_{k+T}) | \mathcal{F}_k] - V(\tilde{z}_k) \leq -\tilde{\varphi}(\tilde{z}_k) \leq 0$. By taking the expectation on both sides of this inequality, we obtain

$$\mathbb{E}[V(\tilde{z}_{k+T})] - \mathbb{E}V(\tilde{z}_k) \leq -\mathbb{E}\tilde{\varphi}(\tilde{z}_k), k \in \mathbb{N}_0. \quad (5)$$

For any $k \in \mathbb{N}$, there is a pair $p, q \in \mathbb{N}_0$ such that $k = pT + q$. It follows from (5) that

$$\begin{aligned} \mathbb{E}[V(\tilde{z}_{pT+q})] - \mathbb{E}V(\tilde{z}_{(p-1)T+q}) &\leq -\mathbb{E}\tilde{\varphi}(\tilde{z}_{(p-1)T+q}), \\ &\quad j = 0, \dots, q; \\ \mathbb{E}[V(\tilde{z}_{iT+m})] - \mathbb{E}V(\tilde{z}_{(i-1)T+m}) &\leq -\mathbb{E}\tilde{\varphi}(\tilde{z}_{(i-1)T+m}), \\ &\quad i = 1, \dots, p-1, m = 0, \dots, T-1 \end{aligned}$$

By summing up all the left and right sides of these inequalities respectively for all the i, j and m , we have

$$\begin{aligned} \sum_{m=0}^{T-1} \left(\mathbb{E}[V(\tilde{z}_{(p-1)T+m}) - \mathbb{E}V(\tilde{z}_m)] \right) + \sum_{j=1}^q \left(\mathbb{E}[V(\tilde{z}_{pT+j}) - \right. \\ \left. \mathbb{E}V(\tilde{z}_{(p-1)T+j}) \right] \leq - \sum_{i=1}^{k-T} \mathbb{E}\tilde{\varphi}(\tilde{z}_i). \quad (6) \end{aligned}$$

As $V(x)$ is nonnegative for all x , from (5) it is easy to observe that the left side of (6) is greater than $-\infty$ even when $k \rightarrow \infty$ since T and q are finite numbers, which implies that $\sum_{i=0}^{\infty} \mathbb{E}\tilde{\varphi}(\tilde{z}_k) < \infty$. By Lemma 4, one knows that $\tilde{\varphi}(\tilde{z}_k) \xrightarrow{a.s.} 0$ as $k \rightarrow \infty$. For $\omega \in \bar{\Omega}$, one can observe that $\tilde{\varphi}(x_k(\omega)) = \varphi(x_k(\omega))$ and $\tilde{z}_k(\omega) = x_k(\omega)$ according to the definitions of $\tilde{\varphi}$ and $\{\tilde{z}_k\}$, respectively. Therefore, $\tilde{\varphi}(\tilde{z}_k(\omega)) = \varphi(x_k(\omega))$ for all $\omega \in \bar{\Omega}$, and subsequently

$$\Pr[\varphi(x_k) \rightarrow 0 | \bar{\Omega}] = \Pr[\tilde{\varphi}(\tilde{z}_k) \rightarrow 0 | \bar{\Omega}] = 1.$$

From the continuity of $\varphi(x)$ it can be seen that $\Pr[x_k \rightarrow \mathcal{D}_1 | \bar{\Omega}] = 1$. The proof of i) is complete since (4) means that the sample paths stay in \mathcal{Q}_λ with probability at least $1 - V(x_0)/\lambda$.

Next, we prove ii) in two steps. We first prove that the origin $x = 0$ is stable in probability. The inequalities $h_1(\|s\|) \leq V(s) \leq h_2(\|s\|)$ imply that $V(x) = 0$ if and only if $x = 0$. Moreover, it follows from $h_1(\|s\|) \leq V(s)$ and the inequality (4) that for any initial condition $x_0 \in \mathcal{Q}_\lambda$,

$$\Pr \left[\sup_{k \in \mathbb{N}} h_1(\|x_k\|) \geq \lambda_1 \right] \leq \Pr \left[\sup_{k \in \mathbb{N}} V(x_k) \geq \lambda_1 \right] \leq \frac{V(x_0)}{\lambda_1}$$

for any $\lambda_1 > 0$. Since h_1 is a class \mathcal{K} function and thus invertible, it can be observed that $\Pr[\sup_{k \in \mathbb{N}} \|x_k\| \geq h_1^{-1}(\lambda)] \leq$

$V(x_0)/\lambda \leq h_2(\|x_0\|)/\lambda$. Then for any $\varepsilon > 0$, it holds that $\lim_{x_0 \rightarrow 0} \Pr[\sup_{k \in \mathbb{N}} \|x_k\| > \varepsilon] \leq \Pr[\sup_{k \in \mathbb{N}} \|x_k\| \geq \varepsilon] = 0$, which means that the origin is stable in probability.

Second, we show the probability that $x_k \rightarrow 0$ tends to 1 as $x_0 \rightarrow 0$. One knows that $\mathcal{D}_1 = \{0\}$ since φ is positive definite in \mathcal{Q}_λ . From i) one knows that x_k converges to $x = 0$ with probability at least $1 - V(x_0)/\lambda$. Since $V(x) \rightarrow 0$ as $x_0 \rightarrow 0$, it holds that $\lim_{x_0 \rightarrow 0} \Pr[\lim_{k \rightarrow \infty} \|x_k\| = 0] \rightarrow 1$. The proof is complete. \square

Particularly, if \mathcal{Q}_λ is positively invariant, i.e., starting from $x_0 \in \mathcal{Q}_\lambda$ all sample paths x_k will stay in \mathcal{Q}_λ for all $k \geq 0$, this corollary follows from Theorem 1 straightforwardly.

Corollary 1. *If \mathcal{Q}_λ is positively invariant w.r.t the system (1) and the assumptions a) and b) in Theorem 1 are satisfied, then the following statements apply:*

- i) *for any initial condition $x_0 \in \mathcal{Q}_\lambda$, x_k converges to \mathcal{D}_1 with probability one;*
- ii) *if moreover $\varphi(x)$ is positive definite on \mathcal{Q}_λ , and $h_1(\|s\|) \leq V(s) \leq h_2(\|s\|)$ for two class \mathcal{K} functions h_1 and h_2 , then $x = 0$ is locally a.s. asymptotically stable in \mathcal{Q}_λ . Furthermore, if $\mathcal{Q}_\lambda = \mathbb{R}^n$, then $x = 0$ is globally a.s. asymptotically stable.*

The next theorem provides a new criterion for exponential convergence and stability of stochastic systems, relaxing the conditions required by Lemma 2.

Theorem 2. *Suppose the assumptions a) and b) of Theorem 1 are satisfied with the inequality of b) strengthened to*

$$\mathbb{E}[V(x_{k+T}) | \mathcal{F}_k] - V(x_k) \leq -\alpha V(x_k), \quad 0 < \alpha < 1. \quad (7)$$

Then the following statements apply:

- i) *for any given $x_0 \in \mathcal{Q}_\lambda$, $V(x_k)$ converges to 0 exponentially at a rate no slower than $(1 - \alpha)^{1/T}$, and x_k converges to $\mathcal{D}_2 := \{x \in \mathcal{Q}_\lambda : V(x) = 0\}$, with probability at least $1 - V(x_0)/\lambda$;*
- ii) *if moreover V satisfies that $c_1\|x\|^a \leq V(x) \leq c_2\|x\|^a$ for some $c_1, c_2, a > 0$, then $x = 0$ is exponentially stable in probability.*

Proof. We first prove i). From the proof of Theorem 1, we know that the sample paths x_k stay in \mathcal{Q}_λ with probability at least $1 - V(x_0)/\lambda$ for any initial condition $x_0 \in \mathcal{Q}_\lambda$ if the assumption a) is satisfied. We next show that for any sample path that always stays in \mathcal{Q}_λ , $V(x_k)$ converges to 0 exponentially fast. Towards this end, we define a random process $\{\hat{z}_k\}$. Let J be as defined in the proof of Theorem 1. If J exists, when $J > T$, let

$$\hat{z}_k = x_k, k < J - T, \quad \hat{z}_k = \varepsilon, k \geq J - T,$$

where ε satisfies $V(\varepsilon) = 0$, when $J \leq T$, let $\hat{z}_k = \varepsilon$ for any $k \in \mathbb{N}_0$; if J does not exist, we let $\hat{z}_k = x_k$ for all $k \in \mathbb{N}_0$.

If the inequality (7) is satisfied, one has $\mathbb{E}[V(\hat{z}_{k+T}) | \mathcal{F}_k] - V(\hat{z}_k) \leq -\alpha V(\hat{z}_k)$. Using this inequality, we next show that $V(\hat{z}_{k+T})$ converges to 0 exponentially. To this end, define a subsequence $Y_m^{(r)} := V(\hat{z}_{mT+r})$, $m \in \mathbb{N}_0$, for each $0 \leq r \leq T-1$. Let $\mathcal{G}_m^{(r)} := \sigma(Y_0^{(r)}, Y_1^{(r)}, \dots, Y_m^{(r)})$, and one knows that $\mathcal{G}_m^{(r)}$ is determined if we know \mathcal{F}_{mT+r} . It then follows from the

inequality (7) that for any r , $\mathbb{E}[Y_{m+1}^{(r)} | \mathcal{G}_m^{(r)}] - Y_m^{(r)} \leq -\alpha Y_m^{(r)}$. We observe from this inequality that

$$\mathbb{E} \left[(1 - \alpha)^{-(m+1)} Y_{m+1}^{(r)} | \mathcal{G}_m^{(r)} \right] - (1 - \alpha)^{-m} Y_m^{(r)} \leq 0.$$

This means that $(1 - \alpha)^{-m} Y_m$ is a supermartingale, and thus there is a finite random number $\bar{Y}^{(r)}$ such that $(1 - \alpha)^{-m} Y_m^{(r)} \xrightarrow{a.s.} \bar{Y}^{(r)}$ for any r . Let $\gamma = \sqrt[1/(1 - \alpha)]{1}$, and then by the definition of $Y_m^{(r)}$ we have $\gamma^{mT} V(\hat{z}_{mT+r}) \xrightarrow{a.s.} \bar{Y}^{(r)}$. Straightforwardly, $\gamma^{mT+r} V(\hat{z}_{mT+r}) \xrightarrow{a.s.} \gamma^r \bar{Y}^{(r)}$. Let $k = mT + r$, $\bar{Y} = \max_r \{\gamma^r \bar{Y}^{(r)}\}$, then it almost surely holds that $\lim_{k \rightarrow \infty} \gamma^k V(\hat{z}_k) \leq \bar{Y}$. From Definition 1, one concludes that $V(\hat{z}_k)$ almost surely converges to 0 exponentially no slower than $\gamma^{-1} = (1 - \alpha)^{1/T}$. From the definition of \hat{z}_k , we know that $V(\hat{z}_k(\omega)) = V(x_k(\omega))$ for all $\omega \in \bar{\Omega}$, with $\bar{\Omega}$ defined in the proof of Theorem 1. Consequently, it holds that

$$\begin{aligned} \Pr \left[\lim_{k \rightarrow \infty} \gamma^k V(x_k) \leq \bar{Y} | \bar{\Omega} \right] \\ = \Pr \left[\lim_{k \rightarrow \infty} \gamma^k V(\hat{z}_k) \leq \bar{Y} | \bar{\Omega} \right] = 1. \end{aligned} \quad (8)$$

The proof of i) is complete since the sample paths stay in \mathcal{Q}_λ with probability at least $1 - V(x_0)/\lambda$.

Next, we prove ii). If the inequalities $c_1 \|x\|^a \leq V(x) \leq c_2 \|x\|^a$ are satisfied, and then we know that $V(x) = 0$ if and only if $x = 0$. Moreover, it follows from (8) that for all the sample paths that stay in \mathcal{Q}_λ it holds that $c_1 \gamma^k \|x\|^a \leq \gamma^k V(x_k) \leq \bar{Y}$ since $c_1 \|x_k\|^a \leq V(x_k)$. Hence, $\|x_k(\omega)\| \leq (\bar{Y}/c_1)^{1/a} \gamma^{-k/a}$ for any $\omega \in \bar{\Omega}$, and one can check that this inequality holds with probability at least $1 - V(x_0)/\lambda$. If $x_0 \rightarrow 0$, we know that $1 - V(x_0)/\lambda \rightarrow 1$, which completes the proof. \square

If \mathcal{Q}_λ is positively invariant, the following corollary follows straightforwardly.

Corollary 2. *If \mathcal{Q}_λ is positively invariant w.r.t the system (1) and suppose the assumptions a) and b) of Theorem 1 are satisfied with the inequality of b) strengthened to (7), the following statements apply:*

- i) *for any given $x_0 \in \mathcal{Q}_\lambda$, $V(x_k)$ converges to 0 exponentially no slower than $(1 - \alpha)^{1/T}$ with probability one;*
- ii) *if moreover V satisfies that $c_1 \|x\|^a \leq V(x) \leq c_2 \|x\|^a$ for some $c_1, c_2, a > 0$, then $x = 0$ is locally a.s. exponentially stable in \mathcal{Q}_λ . Furthermore, if $\mathcal{Q}_\lambda = \mathbb{R}^n$, then $x = 0$ is globally a.s. exponentially stable.*

The following corollary, which can be proven following the same lines as Theorems 1 and 2, shares some similarities to LaSalle's theorem for deterministic systems. It is worth mentioning that the function V here does not have to be radially unbounded.

Corollary 3. *Let $\mathbb{D} \subset \mathbb{R}^n$ be a compact set that is positively invariant w.r.t the system (1). Let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous nonnegative function, and $\mathcal{Q}_\lambda := \{x \in \mathbb{D} : V(x) < \lambda\}$ for some $\lambda > 0$. Assume that $\mathbb{E}[V(x_{k+1}) | \mathcal{F}_k] - V(x_k) \leq 0$ for all k such that $x_k \in \mathcal{Q}_\lambda$, then*

- i) *if there is an integer $T \geq 1$, independent of ω , such that for any $k \in \mathbb{N}_0$, $\mathbb{E}[V(x_{k+T}) | \mathcal{F}_k] - V(x_k) \leq -\varphi(x_k)$, where $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous and satisfies $\varphi(x) \geq 0$ for any*

$x \in \mathcal{Q}_\lambda$, then for any initial condition $x_0 \in \mathcal{Q}_\lambda$, x_k converges to $\bar{\mathcal{D}}_1 := \{x \in \mathcal{Q}_\lambda : \varphi(x) = 0\}$ with probability at least $1 - V(x_0)/\lambda$;

ii) *if the inequality in a) is strengthened to $\mathbb{E}[V(x_{k+T}) | \mathcal{F}_k] - V(x_k) \leq -\alpha V(x_k)$ for some $0 < \alpha < 1$, then for any given $x_0 \in \mathcal{Q}_\lambda$, $V(x_k)$ converges to 0 exponentially at a rate no slower than $(1 - \alpha)^{1/T}$, and x_k converges to $\bar{\mathcal{D}}_2 := \{x \in \mathcal{Q}_\lambda : V(x) = 0\}$, with probability at least $1 - V(x_0)/\lambda$;*

iii) *if \mathcal{Q}_λ is positively invariant w.r.t the system (1), then all the convergence in both i) and ii) takes place almost surely.*

Example 1 Cont. Now let us look back at Example 1 and still choose $V(x) = \|x\|_\infty$ as a stochastic Lyapunov function candidate. It is easy to see that $V(x)$ is a nonnegative supermartingale. To show the stochastic convergence, let $T = 2$ and one can calculate the conditional expectations

$$\begin{aligned} \mathbb{E}[V(x_{k+T}) | x_k, y_k = 1] - V(x_k) \\ = 0.5 \left\| \begin{bmatrix} 0.2x_k^1 \\ 0.8x_k^2 \end{bmatrix} \right\|_\infty + 0.5 \left\| \begin{bmatrix} 0.2x_k^1 \\ 0.6x_k^2 \end{bmatrix} \right\|_\infty - \left\| \begin{bmatrix} x_k^1 \\ x_k^2 \end{bmatrix} \right\|_\infty \\ \leq -0.3V(x_k), \forall x_k \in \mathbb{R}^2. \end{aligned}$$

When $y_k = 2, 3$, it analogously holds that

$$\mathbb{E}[V(x_{k+T}) | x_k, y_k] - V(x_k) \leq -0.3V(x_k), \forall x_k \in \mathbb{R}^2.$$

From these three inequalities one can observe that starting from any initial condition x_0 , $\mathbb{E}V(x)$ decreases at an exponential speed after every two steps before it reaches 0. By Corollary 2, one knows that origin is globally a.s. exponentially stable, consistent with our conjecture. \triangle

Remark 1. *Kushner and other researchers have used more restricted conditions to construct Lyapunov functions than those appearing in our results to analyze asymptotic or exponential stability of random processes [2]–[4]. It is required that $\mathbb{E}[V(x_k)]$ decreases strictly at every step, until $V(x_k)$ reaches a limit value. However, in our result, this requirement is relaxed. In addition, Kushner's results rely on the assumption that the underlying random process is Markovian, but we work with more general random processes.*

In the following sections, we will show how the new Lyapunov criteria can be applied to distributed computation.

III. PRODUCTS OF RANDOM SEQUENCES OF STOCHASTIC MATRICES

In this section, we study the convergence of products of stochastic matrices, where the obtained results on finite-step Lyapunov functions are used for analysis. Let $\Omega_0 := \{1, 2, \dots, m\}$ be the state space and $\mathcal{M} := \{F_1, F_2, \dots, F_m\}$ be the set of m stochastic matrices $F_i \in \mathbb{R}^{n \times n}$. Consider a random sequence $\{W_\omega(k) : k \in \mathbb{N}\}$ on the probability space $(\Omega, \mathcal{F}, \Pr)$, where Ω is the collection of all infinite sequences $\omega = (\omega_1, \omega_2, \dots)$ with $\omega_k \in \Omega_0$, and we define $W_\omega(k) := F_{\omega_k}$. For notational simplicity, we denote $W_\omega(k)$ by $W(k)$. For the backward product of stochastic matrices

$$W(t+k, t) = W(t+k) \cdots W(t+1), \quad (9)$$

where $k \in \mathbb{N}, t \in \mathbb{N}_0$, we are interested in establishing conditions on $\{W(k)\}$, under which it holds that

$\lim_{k \rightarrow \infty} W(k, 0) = L$ for a random matrix $L = \mathbf{1}\xi^\top$ where $\xi \in \mathbb{R}^n$ satisfies $\xi^\top \mathbf{1} = 1$.

Before proceeding, let us introduce some concepts in probability. Let $\mathcal{F}_k = \sigma(W(1), \dots, W(k))$, so that evidently $\{\mathcal{F}_k\}$, $k = 1, 2, \dots$, is an increasing sequence of σ -fields. Let $\phi : \Omega \rightarrow \Omega$ be the shift operator, i.e., $\phi(\omega_1, \omega_2, \dots) = (\omega_2, \omega_3, \dots)$. A random sequence of stochastic matrices $\{W(1), W(2), \dots, W(k), \dots\}$ is said to be *stationary* if the shift operator is measure-preserving. In other words, the sequences $\{W(k_1), W(k_2), \dots, W(k_r)\}$ and $\{W(k_1 + \tau), W(k_2 + \tau), \dots, W(k_r + \tau)\}$ have the same joint distribution for all k_1, k_2, \dots, k_r and $\tau \in \mathbb{N}$. Moreover, a sequence is said to be *stationary ergodic* if it is stationary, and every invariant set \mathcal{B} is trivial, i.e., for every $A \in \mathcal{B}$, $\Pr[A] \in \{0, 1\}$. Here by an invariant set \mathcal{B} , we mean $\phi^{-1}\mathcal{B} = \mathcal{B}$.

A. Convergence Results

We first introduce three classes of stochastic matrices, denoted by $\mathcal{M}_1, \mathcal{M}_2$, and \mathcal{M}_3 , respectively. We say $A \in \mathcal{M}_1$ if A is indecomposable, and aperiodic (such stochastic matrices are also referred to as SIA for short); $A \in \mathcal{M}_2$ if A is scrambling, i.e., no two rows of A are orthogonal; and $A \in \mathcal{M}_3$ if A is Markov, i.e., there exists a column of A such that all entries in this column are positive [37, Ch. 4].

Coefficients of ergodicity serve as a fundamental tool in analyzing the convergence of products of stochastic matrices. In this paper, we employ a standard one. For a stochastic matrix $A \in \mathbb{R}^{n \times n}$, the coefficient of ergodicity $\tau(A)$ is defined by

$$\tau(A) = 1 - \min_{i,j} \sum_{s=1}^n \min(a_{is}, a_{js}). \quad (10)$$

It is known that this coefficient of ergodicity satisfies $0 \leq \tau(A) \leq 1$, and $\tau(A)$ is proper since $\tau(A) = 0$ if and only if all the rows of A are identical. Importantly, it holds that

$$\tau(A) < 1 \quad (11)$$

if and only if $A \in \mathcal{M}_2$ (see [37, p.82]). For any two stochastic matrices A, B , the following property will be critical for the proof in Appendix A:

$$\tau(AB) \leq \tau(A)\tau(B). \quad (12)$$

Before providing our first results in this subsection, we make the following assumption for the sequence $\{W(k)\}$.

Assumption 1. Suppose the sequence of stochastic matrices $\{W(k) : k \in \mathbb{N}\}$ is driven by a random process satisfying the following conditions:

- a) There exists an integer $h > 0$ such that for any $k \in \mathbb{N}_0$, it holds that

$$\Pr[W(k+h, k) \in \mathcal{M}_2] > 0, \quad (13)$$

$$\sum_{i=1}^{\infty} \Pr[W(k+ih, k+(i-1)h) \in \mathcal{M}_2] = \infty; \quad (14)$$

- b) There is a positive number α such that for any $i, j \in \mathbb{N}$, $k \in \mathbb{N}_0$, it holds that $W_{ij}(k) \geq \alpha$ if $W_{ij}(k) > 0$.

Now we are ready to provide our main result on the convergence of stochastic matrices' products.

Theorem 3. Under Assumption 1, the product of the random sequence of stochastic matrices $W(k, 0)$ converges to a random matrix $L = \mathbf{1}\xi^\top$ almost surely.

To prove Theorem 3, consider the stochastic discrete-time dynamical system described by

$$x_{k+1} = F_{y(k+1)} x_k := W(k+1)x_k, k \in \mathbb{N}_0 \quad (15)$$

where $x_k \in \mathbb{R}^n$; the initial state x_0 is a constant with probability one; $y(k) \in \{1, \dots, m\}$ is regarded as the randomly switching signal; and $\{W(1), W(2), \dots\}$ is the random process of stochastic matrices we are interested in. One knows that x_k is adapted to \mathcal{F}_k . Thus, to investigate the limiting behavior of the product (9), it is sufficient to study the limiting behavior of system dynamics (15). We say the state of system (15) reaches an *agreement* state if $\lim_{k \rightarrow \infty} x_k = \mathbf{1}\xi$ for some $\xi \in \mathbb{R}$. Then the agreement of system (15) for any initial state x_0 implies that $W(k, 0)$ converges to a rank-one matrix as $k \rightarrow \infty$ [26].

To investigate the agreement problem, we define $[x_k] := \max_{i \in \mathbb{N}} x_k^i$, $\lfloor x_k \rfloor := \min_{i \in \mathbb{N}} x_k^i$ for any $k \in \mathbb{N}_0$, and

$$v_k = [x_k] - \lfloor x_k \rfloor. \quad (16)$$

For any $k \in \mathbb{N}$, v_k is adapted to \mathcal{F}_k since x_k is. The agreement is said to be reached asymptotically almost surely if $v_k \xrightarrow{a.s.} 0$ as $k \rightarrow \infty$; and it is said to be reached exponentially almost surely with convergence rate no slower than γ^{-1} for some $\gamma > 1$ if $\gamma^k v_k \xrightarrow{a.s.} y$ for some finite $y \geq 0$. The random variable v_k has some important properties given by the following proposition.

Proposition 1. Consider a system $x_{k+1} = Ax_k, k \in \mathbb{N}_0$, where A is a stochastic matrix. For v_k defined in (16), it follows that $v_{k+1} \leq v_k$, and the strict inequality holds for any $x_k \notin \text{span}(\mathbf{1})$ if and only if A is scrambling.

Proof. It is shown in [37] that $v_{k+1} \leq \tau(A)v_k$ with $\tau(\cdot)$ defined in (10). Therefore, the sufficiency follows from (11) straightforwardly. We then prove the necessity by contradiction. Suppose A is not scrambling, and then there must exist at least two rows, denoted by i, j , that are orthogonal. Define the two sets $\mathbf{i} := \{l : a_{il} > 0, l \in \mathbb{N}\}$ and $\mathbf{j} := \{m : a_{jm} > 0, m \in \mathbb{N}\}$, respectively. It follows then from the scrambling property that $\mathbf{i} \cap \mathbf{j} = \emptyset$. Let $x_k^q = 1$ for all $q \in \mathbf{i}$, $x_k^q = 0$ for all $q \in \mathbf{j}$, and let x_k^m be any arbitrary positive number less than 1 for all $m \in \mathbb{N} \setminus (\mathbf{i} \cup \mathbf{j})$ if $\mathbb{N} \setminus (\mathbf{i} \cup \mathbf{j})$ is not empty. Then the states of i and j at time $k+1$ become

$$\begin{aligned} x_{k+1}^i &= \sum_{l=1}^n a_{il} x_k^l = \sum_{l \in \mathbf{i}} a_{il} x_k^l = 1, \\ x_{k+1}^j &= \sum_{l=1}^n a_{jl} x_k^l = \sum_{l \in \mathbf{j}} a_{jl} x_k^l = 0, \end{aligned}$$

and $0 \leq x_{k+1}^m \leq 1$ for all $m \in \mathbb{N} \setminus (\mathbf{i} \cup \mathbf{j})$. This results in $v_{k+1} = v_k = 1$. By contradiction one knows that a scrambling A is necessary for $v_{k+1} < v_k$, which completes the proof. \square

In order to prove Theorem 3, we obtain the following intermediate result.

Proposition 2. For any scrambling matrix $A \in \mathbb{R}^{n \times n}$, the coefficient of ergodicity $\tau(A)$ defined in (10) satisfies

$$\tau(A) \leq 1 - \gamma$$

if all the positive elements of A are lower bounded by $\gamma > 0$.

Proof: Consider any two rows of A , denoted by i, j . Define two sets, $\mathbf{i} := \{s : a_{is} > 0\}$ and $\mathbf{j} := \{s : a_{js} > 0\}$. From the scrambling hypothesis, one knows that $\mathbf{i} \cap \mathbf{j} \neq \emptyset$. Thus it holds that

$$\sum_{s=1}^n \min(a_{is}, a_{js}) = \sum_{s \in \mathbf{i} \cap \mathbf{j}} \min(a_{is}, a_{js}) \geq \gamma.$$

Then from the definition of $\tau(A)$, it is easy to see

$$\tau(A) = 1 - \min_{i,j} \sum_{s=1}^n \min(a_{is}, a_{js}) \leq 1 - \gamma,$$

which completes the proof. \blacksquare

We are in the position to prove Theorem 3 by showing that $v_k \xrightarrow{a.s.} 0$ as $k \rightarrow \infty$, where the results obtained in Corollary 3 will be used.

Proof of Theorem 3: Let $V(x_k) = v_k$ be a finite-step stochastic Lyapunov function candidate for the system dynamics (15). It is easy to see $V(x) = 0$ if and only if $x \in \text{span}(\mathbf{1})$. Since all $W(k)$ are stochastic matrices, we observe that $\mathbb{E}[V(x_{k+1}) | \mathcal{F}_k] - V(x_k) \leq 0$ from Proposition 1, which implies that $V(x_k)$ is exactly a supermartingale with respect to \mathcal{F}_k . From Lemma 3, we know $V(x_k) \xrightarrow{a.s.} \bar{V}$ for some \bar{V} because $V(x_k) \geq 0$ and $\mathbb{E}V(x_k) < \infty$. From Assumption 1, we know that there is an h such that the product $W(k+h, k)$ is scrambling with positive probability for any k . Let \mathcal{W}_k be the set of all possible $W(k+h, k)$ at time k , and n_k the cardinality of \mathcal{W}_k . Let n_k^s be the number of scrambling matrices in \mathcal{W}_k . We denote each of these scrambling matrices and each of non-scrambling matrices by $S_k^i, i = 1, \dots, n_k^s$ and $\bar{S}_k^j, j = 1, \dots, n_k - n_k^s$, respectively. The probabilities of all the possible $W(k+h, k)$ sum to 1, i.e.,

$$\sum_{i=1}^{n_k^s} \Pr[S_k^i] + \sum_{j=1}^{n_k - n_k^s} \Pr[\bar{S}_k^j] = 1. \quad (17)$$

Then the conditional expectation of $V(x)$ after finite steps for any k becomes

$$\begin{aligned} \mathbb{E}[V(x_{k+h}) | \mathcal{F}_k] - V(x_k) &= \mathbb{E}[V(W(k+h, k)x_k)] - V(x_k) \\ &\leq \mathbb{E}[\tau(W(k+h, k))] V(x_k) - V(x_k), \end{aligned}$$

where $\tau(\cdot)$ is given by (10). One can calculate that

$$\begin{aligned} &\mathbb{E}[\tau(W(k+h, k))] - 1 \\ &= \sum_{i=1}^{n_k^s} \Pr[S_k^i] \tau(S_k^i) + \sum_{j=1}^{n_k - n_k^s} \Pr[\bar{S}_k^j] \tau(\bar{S}_k^j) - 1 \\ &\leq \sum_{i=1}^{n_k^s} \Pr[S_k^i] (\tau(S_k^i) - 1), \end{aligned}$$

where Proposition 1 and equation (17) have been used. From Assumption 1.b), we know that the positive elements of $W(k)$ are lower bounded by α , and thus the positive elements of

S_k^i in (18) are lower bounded by α^h . Thus $\tau(S_k^i) \leq 1 - \alpha^h$ according to Proposition 2, and it follows that

$$\begin{aligned} &\mathbb{E}[V(x_{k+h}) | \mathcal{F}_k] - V(x_k) \\ &\leq - \sum_{i=1}^{n_k^s} \Pr[S_k^i] \alpha^h \mathbb{E}V(x_k) := \varphi_k(x_k). \end{aligned} \quad (18)$$

By iterating, one can easily show that

$$\begin{aligned} \mathbb{E}[V(x_{nh})] - V(x_0) &\leq - \sum_{k=0}^{n-1} \varphi_k(x_k) \\ &= - \sum_{k=0}^{n-1} \sum_{i=1}^{n_k^s} \Pr[S_k^i] \alpha^h \mathbb{E}V(x_k). \end{aligned} \quad (19)$$

It then follows that $V(x_0) - \mathbb{E}[V(x_{nh})] < \infty$ even when $n \rightarrow \infty$, since $V(x) \geq 0$. According to the condition (14), we know $\sum_{k=0}^{n-1} \sum_{i=1}^{n_k^s} \Pr[S_k^i] = \infty$. By contradiction, it is easy to infer that $\mathbb{E}V(x_k) \xrightarrow{a.s.} 0$. Since we have already shown that $V(x_k) \xrightarrow{a.s.} \bar{V}$ for some random $\bar{V} \geq 0$, one can conclude that $V(x_k) \xrightarrow{a.s.} 0$. For any given $x_0 \in \mathbb{R}^n$, define the compact set $\mathcal{Q} := \{x : \lceil x \rceil \leq \lceil x_0 \rceil, \lfloor x \rfloor \geq \lfloor x_0 \rfloor\}$. For any random sequence $\{W(k)\}$, it follows from the system dynamics (15) that

$$\begin{aligned} \lceil x_k \rceil &\leq \lceil x_{k-1} \rceil \leq \dots \leq \lceil x_1 \rceil \leq \lceil x_0 \rceil, \\ \lfloor x_k \rfloor &\geq \lfloor x_{k-1} \rfloor \geq \dots \geq \lfloor x_1 \rfloor \geq \lfloor x_0 \rfloor, \end{aligned}$$

and thus x_k will remain within \mathcal{Q} . From Corollary 3, we know that x_k asymptotically converges to $\{x \in \mathcal{Q} : \varphi_k(x) = 0\}$, or equivalently, $\{x \in \mathcal{Q} : V(x) = 0\}$ almost surely as $k \rightarrow \infty$ since $V(x)$ is continuous. In other words, for any $x_0 \in \mathbb{R}^n$, $x_k \xrightarrow{a.s.} \zeta \mathbf{1}$ for some $\zeta \in \mathbb{R}$, which proves Theorem 3. \blacksquare

Compared to the existing results, Theorem 3 has provided a quite relaxed condition for the convergence of the backward product (9) determined by the random sequence $\{W(k)\}$ to a rank-one matrix: over any time interval of length h , i.e., $[k, k+h]$ for any $k \in \mathbb{N}_0$, the product $W(k+h) \cdots W(k+1)$ has positive probability to be scrambling. The following corollary follows straightforwardly since a Markov matrix is certainly scrambling.

Corollary 4. For a random sequence $\{W(k) : k \in \mathbb{N}\}$, the product (9) converges to a random matrix $L = \mathbf{1}\xi^\top$ almost surely if there exists an integer h such that for any k the product $W(k+h, k)$ is a Markov matrix with positive probability and $\sum_{i=1}^{\infty} \Pr[W(k+ih, k + (i-1)h) \in \mathcal{M}_3] = \infty$.

Next we assume that the sequence $\{W(k)\}$ is driven by an underlying stationary process. Then the condition in Theorem 3 can be further relaxed. Let us make the following assumption and provide another theorem in this subsection.

Assumption 2. Suppose the random sequence of stochastic matrices $\{W(k) : k \in \mathbb{N}\}$ is driven by a stationary process satisfying the following conditions:

a) There exists an integer $h > 0$ such that for any $k \in \mathbb{N}_0$, it holds that

$$\Pr[W(k+h, k) \in \mathcal{M}_1] > 0; \quad (20)$$

b) There is a positive number α such that for any $i, j \in \mathbb{N}, k \in \mathbb{N}_0$, it holds that $W_{ij}(k) \geq \alpha$ if $W_{ij}(k) > 0$.

In other words, we suppose in Assumption 2 that any corresponding matrix product of length h becomes an SIA

matrix with positive probability, and the positive elements for any matrix in \mathcal{M} are uniformly lower bounded away from some positive value.

Theorem 4. *Under Assumption 2, the product of the random sequence of stochastic matrices $W(k, 0)$ converges to a random matrix $L = \mathbf{1}\xi^\top$ almost surely.*

If two stochastic matrices A_1 and A_2 have zero elements in the same positions, we say these two matrices are of the same type, denoted by $A_1 \sim A_2$. Obviously, it holds the trivial case $A_1 \sim A_1$. One knows that for any SIA matrix A , there exists an integer l such that A^l is scrambling; it is easy to extend this to the inhomogeneous case, i.e., any product of l stochastic matrices of the same type of A is scrambling if all the matrices' elements are lower bounded away by some positive number. We are now ready to prove Theorem 4.

Proof of Theorem 4: Since $\{W(k)\}$ is driven by a stationary process, we know that $\{W(t+h), \dots, W(t+1)\}$ has the same joint distribution as $\{W(t+2h), \dots, W(t+h+1)\}$ for any $t \in \mathbb{N}_0, h \in \mathbb{N}$. For the h given in Assumption 2, there exists an SIA matrix A such that $\Pr[W(t+kh+h, t+kh+1) = A] > 0$. Thus it follows that $\Pr[W(t+kh+2h, t+kh+1) = A] > 0$ for any $k \in \mathbb{N}_0$. Thus

$$\Pr \left[\begin{array}{c} W(t+(k+2)h, t+(k+1)h) \\ \sim W(t+(k+1)h, t+kh) \end{array} \middle| W(h, t+kh) \right] > 0.$$

When $W(t+h, t) \in \mathcal{M}_1$, which happens with positive probability, we have

$$\begin{aligned} & \Pr \left[\begin{array}{c} W(t+2h, t+h) \sim W(t+h, t), \\ W(t+h, t) \in \mathcal{M}_1 \end{array} \right] \\ &= \Pr \left[\begin{array}{c} W(t+2h, t+h) \\ \sim W(t+h, t) \end{array} \middle| \Pr[W(t+h, t) \in \mathcal{M}_1] \right] \\ & \quad \cdot \Pr[W(t+h, t) \in \mathcal{M}_1] > 0. \end{aligned}$$

By recursion one can conclude that all the m products $W(t+(k+1)h, t+kh), k \in \{0, \dots, m-1\}$, occur as the same SIA type with positive probability. Since all the products $W(t+(k+1)h, t+kh)$ are of the same type, one can choose m such that $W(t+mh, t)$ is scrambling. This in turn implies that $\Pr[W(t+mh, t) \in \mathcal{M}_2] > 0$, and the property of stationary process makes sure that (14) holds. The conditions in Assumption 1 are therefore all satisfied, and then Theorem 4 follows from Theorem 3. ■

Remark 2. *Theorems 3 and 4 have established some sufficient conditions for the convergence of a random sequence of stochastic matrices to a rank-one matrix. A further question is how these results can be applied to controlling distributed computation processes. To answer this question, let us still consider a finite set of stochastic matrices $\mathcal{M} = \{F_1 \dots, F_m\}$, from which each $W(k)$ in the random sequence $\{W(k)\}$ is sampled. It is defined in [38] that \mathcal{M} is a consensus set if the arbitrary product $\prod_{i=1}^k W(i), W(i) \in \mathcal{M}$, converges to a rank-one matrix. However, it has also been shown that to decide whether \mathcal{M} is a consensus set is an NP-hard problem [38], [39]. For a non-consensus set \mathcal{M} , it is always not obvious how to find a deterministic sequence that converges,*

especially when \mathcal{M} has a large number of elements and F_i has zero diagonal entries. However, the convergence can be ensured almost surely by introducing some randomness in the sequence, provided that there is a convergent deterministic sequence intrinsically.

B. Estimate of Convergence Rate

In Section III-A, we have shown how the product $W(k, 0)$ determined by a random process asymptotically converges to a rank-one matrix W a.s. as $k \rightarrow \infty$. However, the convergence rate for such a randomized product is not yet clear. It is quite challenging to investigate how fast the process converges, especially when each $W(k)$ may have zero diagonal entries. In this subsection, we address this problem by employing finite-step stochastic Lyapunov functions. Now let us present the main result on convergence rate.

Theorem 5. *In addition to Assumption 1, if there exists a number $p, 0 < p < 1$, such that for any $k \in \mathbb{N}_0$*

$$\Pr[W(h, k) \in \mathcal{M}_2] \geq p > 0,$$

then the almost sure convergence of the product $W(k, 0)$ to a random matrix $L = \mathbf{1}\xi^\top$ is exponential, and the rate is no slower than $(1 - p\alpha^h)^{1/h}$.

Proof: Choosing $V(x_k) = v_k$ as a finite-step stochastic Lyapunov function candidate, from (18) we have

$$\begin{aligned} & \mathbb{E}[V(x_{k+h}) | \mathcal{F}_k] - V(x_k) \\ & \leq - \sum_{i=1}^{n_k^s} \Pr[S_k^i] \alpha^h V(x_k). \end{aligned} \quad (21)$$

Furthermore, it is easy to see that

$$\sum_{i=1}^{n_k^s} \Pr[S_k^i] = \Pr[W(h, t) \in \mathcal{M}_2] \geq p,$$

Substituting it into (21) yields

$$\mathbb{E}[V(x_{k+h}) | \mathcal{F}_k] \leq (1 - p\alpha^h) V(x_k).$$

It follows from Corollary 3 that $V(x_{k+h}) \xrightarrow{a.s.} 0$, with an convergence rate no slower than $(1 - p\alpha^h)^{1/h}$. In other words, the agreement is reached exponentially almost surely, which implies Theorem 5. ■

Theorem 5 has established the almost sure exponential convergence rate for the product of $\{W(k)\}$. If any subsequence $\{W(k+1), \dots, W(k+2), W(k+h)\}$ can result in a scrambling product $W(k+h, k)$ with positive probability and this probability is lower bounded away by some positive number, and then the convergence rate is exponential. Interestingly, the greater this lower bound is, the faster the convergence becomes. If we consider a special random sequence which is driven by a stationary ergodic process, the exponential convergence rate follows without any other conditions apart from Assumption 2, and an alternative proof is given in Appendix A.

Corollary 5. *Suppose the random process governing the evolution of the sequence $\{W(k) : k \in \mathbb{N}\}$ is stationary ergodic, then the product $W(k, 0)$ converges to a random rank-one matrix at an exponential rate almost surely under Assumption 2.*

C. Connections to Markov Chains

In this subsection, we show that Theorems 4, and 5 are the generalizations of some well known results for Markov chains in [37], [40]. A fundamental result on inhomogeneous Markov chains is as follows.

Lemma 5 ([37, Th. 4.10], [40]). *If the product $W(k, t)$, formed from a sequence $\{W(k)\}$, satisfies $W(t+k, t) \in \mathcal{M}_1$ for any $k \geq 1, t \geq 0$, and $W_{ij}(k) \geq \alpha$ whenever $W_{ij}(k) > 0$, then $W(k, 0)$ converges to a rank-one matrix as $k \rightarrow \infty$.*

Let h be the number of distinct types of scrambling matrices of order n . It is known that the product $W(k+h, k)$ is scrambling for any k . In this case, we may take the probability of each product $W(k+h, k)$ being scrambling as $p = 1$, and as an immediate consequence of Theorem 5, we know that $W(k, 0)$ converges to a rank-one matrix at an exponential rate that is no slower than $(1 - \alpha^h)^{1/h}$. This convergence rate is consistent with what is estimated in [37, Th. 4.10]. This also applies to the homogeneous case where $W(k) = W$ for any k with W being scrambling. Moreover, it is known that the condition can be relaxed by just requiring W to be SIA to ensure the convergence, which is an immediate consequence of Theorem 4.

In next section, we discuss how the results can be further applied to the context of asynchronous computations.

IV. ASYNCHRONOUS AGREEMENT OVER POSSIBLY PERIODIC NETWORKS

In this section, we take each component x^j in x from (15) as the state of agent i in an n -agent system. Define the distributed coordination algorithm

$$x^i(t_{k+1}) = \sum_{j=1}^n w_{ij} x^j(t_k), k \in \mathbb{N}_0, i \in \mathbb{N}, \quad (22)$$

where the averaging weights $w_{ij} \geq 0$, $\sum_{j=1}^n w_{ij} = 1$, and t_k denote the time instants when updating actions happen. Here we assume the initial state $x(t_0)$ is given. It is always assumed that $T_1 \leq t_{k+1} - t_k \leq T_2$, where $t_0 = 0$ and T_1, T_2 are positive numbers. We say the states of system (22) reach *agreement* if $\lim_{k \rightarrow \infty} x(t_k) = 1\zeta$, mentioned in Section III. Let $W = [w_{ij}] \in \mathbb{R}^{n \times n}$, and obviously W is a stochastic matrix. The algorithm (22) can be rewritten as $x(t_{k+1}) = Wx(t_k)$. In fact, the matrix W can be associated with a directed, weighted graph $\mathcal{G}_W = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} := \{1, 2, \dots, n\}$ is the vertex set and \mathcal{E} is the edge set for which $(i, j) \in \mathcal{E}$ if $w_{ji} > 0$. The graph \mathcal{G}_W is called a *rooted* one if there exists at least one vertex, called a *root*, from which any other vertex can be reached. It is known that agents are able to reach agreement for all $x(0)$ if W is SIA ([37], [40]). However, the situations when W is not SIA have not been studied before, although they appear often in real systems, such as social networks. As we are interested in studying the agreement problem when W is possibly periodic, let us define periodic stochastic matrices.

Definition 4. A stochastic matrix $A \in \mathbb{R}^{n \times n}$ is said to be periodic with period $d > 1$ if d is the common divisor of all the t such that $A^{m+t} \sim A^m$ for a sufficiently large integer m .

Definition 4 is a generalization of the definition of an irreducible periodic matrix [37, Def. 1.6]. In this definition, a periodic stochastic matrix is not necessarily irreducible. With a slight abuse of terminology, we say the graph \mathcal{G}_W is *periodic* if the associated matrix W is periodic.

In the context of distributed computation, it is always assumed that each individual computational unit in the network has access to its own latest state while implementing the iterative update rules [19], [21]. A class of situations that have received considerably less attention in the literature arise when some individuals are not able to obtain their own state, a case which can result from memory loss. Similar phenomena have also been observed in social networks while studying the evolution of opinions. Self-contemptuous people change their opinions solely in response to the opinions of others. The existence of computational units or individuals who are not able to access their own states sometimes might result in the computational failure or opinions' disagreement. As such an example, a periodic matrix W , which must have all zero diagonal entries (no access to their own states for all individuals), always leads the system (22) to oscillation. This is because for a periodic W , W^k never converges to a matrix with identical rows as $k \rightarrow \infty$. Instead, the positions of W^k that have positive values are periodically changing with k , resulting in a periodically changing value of $W^k x(0)$. This motivates us to investigate the particular case where W is possibly periodic.

In this section, we show that agreement can be reached even when W is periodic, *just by introducing asynchronous updating events to the coupled agents*. In fact, perfect synchrony is hard to realize in practice as it is difficult for all agents to have access to a common clock according to which they coordinate their updating actions, while asynchrony is more likely. Researchers have studied how agreement can be preserved with the existence of asynchrony, see e.g., [41], [42]. Unlike these works, we approach the same problem from a different aspect, where agreement occurs just because of asynchrony. A counterpart of this problem where W is irreducible and periodic has been covered in our earlier work [43]. We consider a more general case in this section where W can be reducible.

To proceed, we define a framework of randomly asynchronous updating events. It is usually legitimate to postulate that on occasions more than one, but not all, agents may update. Assume that each agent is equipped with a clock, which need not be synchronized with other clocks. The state of each agent remains unchanged except when an activation event is triggered by its own clock. Denote the set of event times of the i th agent by $\mathcal{T}^i = \{0, t_1^i, \dots, t_k^i, \dots\}$, $k \in \mathbb{N}$. At the event times, agent i updates its state obeying the asynchronous updating rule

$$x_i(t_{k+1}^i) = \sum_{j=1}^n w_{ij} x_j(t_k^i), \quad (23)$$

where $i \in \mathbb{N}$. We assume that the clocks which determine the updating events for the agents are driven by an underlying random process. The following assumption is important for the analysis.

Assumption 3. For any agent i , the intervals between two event times, denoted by $h_k^i = t_k^i - t_{k-1}^i$, are such that

- (i) h_k^i are upper bounded with probability 1 for all k and all i ;
- (ii) $\{h_k^i : k \in \mathbb{N}_0\}$ is a random sequence, with $\{h_k^1\}$, $\{h_k^2\}$, \dots , $\{h_k^n\}$ being mutually independent.

Assumption 3 ensures that an agent can be activated again within finite time after it is activated at t_{k-1}^i for all $k \in \mathbb{N}$, which implies that all agents will update their states for infinitely many times in the long run. In fact, Assumption 3 can be satisfied if the agents are activated by mutually independent Poisson clocks or at rates determined by mutually independent Bernoulli processes ([44, Ch. 6], [32, Ch. 2]).

Let $\mathcal{T} = \{t_0, t_1, t_2, \dots, t_k, \dots\}$ denote all event times of all the n agents, in which the event times have been relabeled in a way such that $t_0 = 0$ and $t_\tau < t_{\tau+1}$, $\tau = \{0, 1, 2, \dots\}$. This idea has been used in [45] and [21] to study asynchronous iterative algorithms. One situation may occur in which there exists some k such that $t_k \in \mathcal{T}^i$ and $t_k \in \mathcal{T}^j$ for some i, j , which implies more than one agent is activated at some event times. Although this is not likely to happen when the underlying process is some special random ones like Poisson, our analysis and results will not be affected. For simplicity, we rewrite the set of event times as $\mathcal{T} = \{0, 1, 2, \dots, k, \dots\}$. Then the system with asynchronous updating can be treated as one with discrete-time dynamics in which the agents are permitted to update only at certain event times $k, k \in \mathbb{N}$, according to the updating rule (23) at each time k . Since each $k \in \mathcal{T}$ can be the event time of any subset of agents, we can associate any set of event times $\{k+1, k+2, \dots, k+h\}$ with the updating sequence of agents $\{\lambda(k+1), \lambda(k+2), \dots, \lambda(k+h)\}$ with $\lambda(i) \in \mathcal{V}$. Under Assumption 3, one knows that this updating sequence can be *arbitrarily ordered*, and each possible sequence can occur with positive probability, though the particular value is not of concern.

Assume at time k , $m \geq 1$ agents are activated, labeled by k_1, k_2, \dots, k_m , then we define the following matrices

$$W(k) = [u_1, \dots, w_{k_1}^\top, u_{k+1}, \dots, w_{k_m}^\top, \dots, u_n]^\top, \quad (24)$$

where $u_i \in \mathbb{R}^n$ is the i th column of the identity matrix I_n and $w_k \in \mathbb{R}^n$ denotes the k th row of W . We call $W(k)$ the *asynchronous updating matrix* at time k . Then the asynchronous updating rule (23) becomes

$$x_k = W(k)x_{k-1}, k \in \mathbb{N}, \quad (25)$$

where $\{W(k)\}$ is a random sequence of asynchronous updating matrices which are stochastic, and $x_0 \in \mathbb{R}^n$ is a given initial state. We say the *asynchronous agreement* is reached if x_k converges to a scaled all-one vector when the agents update asynchronously. It suffices to study the convergence of the product $W(k) \dots W(2)W(1)$ to a rank-one matrix. We now show the asynchronous agreement is reached almost surely even when the graph is periodic. A necessary and sufficient condition for the graph is obtained, under which the agreement can always be reached.

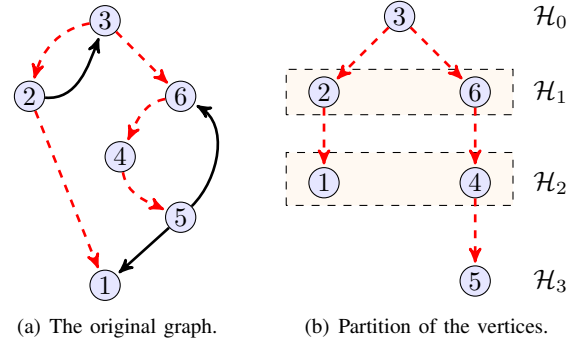


Fig. 1. An illustration of the graph partition; the hierarchical subsets: $\mathcal{H}_0 = \{3\}$, $\mathcal{H}_1 = \{2, 6\}$, $\mathcal{H}_2 = \{1, 4\}$, $\mathcal{H}_3 = \{5\}$; for example, $\{3, 2, 6, 1, 4, 5\}$ is a hierarchical updating vertex sequence.

Theorem 6. Suppose the agents coupled by a network update asynchronously under Assumption 3, then they reach agreement almost surely if and only if the network is rooted, i.e., the matrix W is indecomposable.

To prove this theorem, we need to introduce some additional concepts and results. It is equivalent to say the associated graph \mathcal{G}_W is rooted if W is indecomposable. Denote the set of all the roots of \mathcal{G}_W by $\mathbf{r} \subseteq \mathcal{V}$. We can partition the vertices of \mathcal{G}_W into some hierarchical subsets as follows. For any $\kappa \in \mathbf{r}$, there must exist at least one directed spanning tree rooted at κ , see e.g., Fig. 1 (a). We select any of these directed spanning trees, denoted by \mathcal{G}_W^s . There exists a directed path from κ to any other vertex $i \in \mathcal{V} \setminus \kappa$, see e.g., Fig. 1 (b). Let l_i be the length of the directed path from κ to i , and there exists an integer $L \leq n$ such that $l_i < L$ for all i . Define

$$\mathcal{H}_r := \{i : l_i = r\}, r = 1, \dots, L-1,$$

and $\mathcal{H}_0 = \{\kappa\}$. From this definition, one can partition the vertices of \mathcal{G}_W^s into L hierarchical subsets, i.e., $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_{L-1}$, according to the vertices' distances to the root κ . Let n_r be the number of vertices in the subset \mathcal{H}_r , $0 \leq r \leq L-1$ (see the example in Fig. 1 (b)). Note that given a spanning tree, its corresponding hierarchical subsets \mathcal{H}_r 's are uniquely determined.

Definition 5. An updating vertex sequence of length n is said to be hierarchical if it can be partitioned into some successive subsequences, denoted by $\{\mathcal{A}_0, \dots, \mathcal{A}_{L-1}\}$ with $\mathcal{A}_r = \{\lambda_r(1), \lambda_r(2), \dots, \lambda_r(n_r)\}$, such that $\bigcup_{k=1}^{n_r} \lambda_r(k) = \mathcal{H}_r$ for all $r = 0, \dots, L-1$, where \mathcal{H}_r 's are the hierarchical subsets of some spanning tree \mathcal{G}_W^s in \mathcal{G}_W .

Proposition 3. If agents coupled by \mathcal{G}_W update in a hierarchical sequence $\{a_1, \dots, a_n\}$, $a_i \in \mathcal{V}$ for all i , the product of the corresponding asynchronous updating matrices, $\bar{W} := W_{a_n} \dots W_{a_2}W_{a_1}$, is a Markov matrix.

To prove this proposition, we define an operator $\mathcal{N}(\cdot, \cdot)$ for any stochastic matrix and any subset $\mathcal{S} \in \mathcal{V}$

$$\mathcal{N}(A, \mathcal{S}) := \{j : A_{ij} > 0, i \in \mathcal{S}\},$$

and we write $\mathcal{N}(A, \{i\})$ as $\mathcal{N}(A, i)$ for brevity. It is easy to check then for any two stochastic matrices $A_1, A_2 \in \mathbb{R}^{n \times n}$ and for any subset $\mathcal{S} \in \mathcal{V}$, it holds that

$$\mathcal{N}(A_2 A_1, \mathcal{S}) = \mathcal{N}(A_1, \mathcal{N}(A_2, \mathcal{S})). \quad (26)$$

Proof of Proposition 3: It suffices to show that all $i \in \mathcal{V}$ share at least one common neighbor in the graph $\mathcal{G}_{\tilde{W}}$, i.e.,

$$\bigcap_{i=1}^n \mathcal{N}(\tilde{W}, i) \neq \emptyset. \quad (27)$$

We rewrite the product of asynchronous updating matrices into

$$\tilde{W} = \{W_{\lambda_{L-1}(1)} \cdots W_{\lambda_{L-1}(n_{L-1})} \cdots W_{\lambda_{L-2}(1)} \cdots W_{\lambda_0(1)}\}.$$

For any distinct $i, j \in \mathcal{V}$, we know that $\mathcal{N}(W_j, i) = \{i\}$ from the definition of asynchronous updating matrices. Then for any $\lambda_r(t) \in \mathcal{H}_r, t \in \{1, \dots, n_r\}, r \in \{1, \dots, L-1\}$, it holds that

$$\begin{aligned} & \mathcal{N}(\tilde{W}, \lambda_r(t)) \\ &= \mathcal{N}(W_{\lambda_r(t)} W_{\lambda_r(t+1)} \cdots W_{\lambda_r(n_r)} \cdots W_{\lambda_0(1)}, \lambda_r(t)) \\ &= \mathcal{N}(W_{\lambda_r(t+1)} \cdots W_{\lambda_r(n_r)} \cdots W_{\lambda_0(1)}, \mathcal{N}(W_{\lambda_r(t)}, \lambda_r(t))), \end{aligned}$$

where the property (26) has been used. From Definition 5, one knows that there exists at least one vertex $\lambda_{r-1}(t_1) \in \mathcal{H}_{r-1}$ that can reach $\lambda_r(t)$ in \mathcal{G}_W and subsequently in $\mathcal{G}_{W_{\lambda_r(t)}}$, which implies

$$\lambda_{r-1}(t_1) \in \mathcal{N}(W_{\lambda_r(t)}, \lambda_r(t)).$$

It then follows

$$\begin{aligned} & \mathcal{N}(W_{\lambda_r(t+1)} \cdots W_{\lambda_r(n_r)} \cdots W_{\lambda_0(1)}, \lambda_{r-1}(t_1)) \\ & \subseteq \mathcal{N}(\tilde{W}, \lambda_r(t)). \end{aligned}$$

Similarly, one obtains

$$\begin{aligned} & \mathcal{N}(W_{\lambda_r(t+1)} \cdots W_{\lambda_r(n_r)} \cdots W_{\lambda_0(1)}, \lambda_{r-1}(t_1)) \\ &= \mathcal{N}(W_{\lambda_{r-1}(t_1)} \cdots W_{\lambda_r(n_r)} \cdots W_{\lambda_0(1)}, \lambda_{r-1}(t_1)) \\ &= \mathcal{N}(W_{\lambda_{r-1}(t_1+1)} \cdots W_{\lambda_0(1)}, \mathcal{N}(W_{\lambda_{r-1}(t_1)}, \lambda_{r-1}(t_1))) \\ &\supseteq \mathcal{N}(W_{\lambda_{r-1}(t_1+1)} \cdots W_{\lambda_0(1)}, \lambda_{r-2}(t_2)). \end{aligned}$$

As a recursion, it must be true that

$$\mathcal{N}(W_{\lambda_0(1)}, \kappa) \subseteq \mathcal{N}(\tilde{W}, \lambda_r(t)), \quad (28)$$

where κ is a root of \mathcal{G}_W^s . In fact, it holds that $\lambda_0(1) = \kappa$, and then we know

$$\mathcal{N}(W_{\lambda_0(1)}, \kappa) = \mathcal{N}(W_\kappa, \kappa) = \mathcal{N}(W, \kappa). \quad (29)$$

Substituting (29) into (28) leads to

$$\mathcal{N}(W, \kappa) \subseteq \mathcal{N}(\tilde{W}, \lambda_r(t))$$

for all $\lambda_r(t)$. Since $\bigcup_{r,t} \{\lambda_r(t)\} = \mathcal{V}$, we know

$$\mathcal{N}(W, \kappa) \subseteq \bigcap_{r,t} \mathcal{N}(\tilde{W}, \lambda_r(t)).$$

Straightforwardly, (27) follows, which completes the proof. ■

Since the hierarchical sequences will appear with positive probability in any sequence of length n , one can easily prove the following proposition by letting $l = n$.

Proposition 4. *There exists an integer l such that the product $W(k+l) \cdots W(k+1)$, where $W(k)$ is given in (25), is a Markov matrix with positive probability for any $k \in \mathbb{N}$.*

Proof of Theorem 6: We prove the necessity by contradiction. Suppose the matrix W is decomposable. Then there are at least two sets of vertices that are isolated from each other. Then agreement will never happen between these two isolated groups if they have different initial states. Let $l = n$, in view of Corollary 4, the sufficiency follows directly from Proposition 4, which completes the proof. ■

Remark 3. *Note that the hierarchical sequence is a particular type of updating orders that results in a Markov matrix as the product of the corresponding updating matrices. We have identified another type of updating orders in our earlier work when W is irreducible and periodic [43]. It is of great interest for future work to look for other updating mechanisms to enable the appearance of Markov matrices or scrambling matrices, which plays a crucial role in giving rise to asynchronous agreement.*

In the next section, we look into another application in solving linear algebraic equations.

V. TO SOLVE LINEAR ALGEBRAIC EQUATIONS

Researchers have been quite interested in solving a system of linear algebraic equations in the form of $Ax = b$ in a distributed way [28], [29], [46], [47]. In this section we deal with the problem under the assumption that this system of equations has at least one solution. The set of equations is decomposed into smaller sets and distributed to a network of n processors, referred to as agents, to be solved in parallel. Agents can receive information from their neighbors and the neighbor relationships are described by a time-varying n -vertex directed graph $\mathcal{G}(t)$ with self-arcs. When each agent knows only the pair of real-valued matrices $(A_i^{n_i \times m}, b_i^{n_i \times 1})$, the problem of interest is to devise local algorithms such that all n agents can iteratively compute the same solution to the linear equation $Ax = b$, where $A = [A_1^\top, A_2^\top, \dots, A_n^\top]^\top, b = [b_1^\top, b_2^\top, \dots, b_n^\top]^\top$ and $\sum_{i=1}^n n_i = m$. A distributed algorithm to solve the problem is introduced in [30], where the iterative updating rule for each agent i is described by

$$x_{k+1}^i = x_k^i - \frac{1}{d_k^i} P_i \left(d_k^i x_k^i - \sum_{j \in \mathcal{N}_i(k)} x_k^j \right), k \in \mathbb{N}, \quad (30)$$

where $x_k^i \in \mathbb{R}^m$, d_k^i is the number of neighbors of agent i at time k , $\mathcal{N}_i(k)$ is the collection of i 's neighbors, P_i is the orthogonal projection on the kernel of A_i , and the initial value x_1^i is any solution to the equations of $A_i x = b_i$.

Before proceeding, let us introduce some concepts in graph theory. Given two directed graphs \mathcal{G}_1 and \mathcal{G}_2 with the vertex set \mathcal{V} , the composition of them, denoted by $\mathcal{G}_2 \circ \mathcal{G}_1$, is a directed graph with the vertex set \mathcal{V} and edge set defined in such a way that (i, j) is an arc of the composition just in case there is a vertex i_1 such that (i, i_1) is an edge in \mathcal{G}_1 and meanwhile (i_1, j) is an edge in \mathcal{G}_2 . Given a sequence of graphs $\{\mathcal{G}(1), \mathcal{G}(2), \dots, \mathcal{G}(k)\}$, a *route* over it is a sequence of vertices i_0, i_1, \dots, i_k such that (i_{j-1}, i_j) is an edge in $\mathcal{G}(j)$ for all $1 \leq j \leq k$.

The results in [30] have shown that all x_k^i converge to the same solution exponentially fast if the sequence of graphs $\mathcal{G}(t)$ is repeatedly jointly strongly connected. This condition

requires that for some integer l , the composition of the sequence of graphs, $\{\mathcal{G}(k), \dots, \mathcal{G}(k+l-1)\}$, must be strongly connected for any t . It is not so easy to satisfy this condition if the network is changing randomly. Now assume that the evolution of the sequence of graphs $\{\mathcal{G}(1), \dots, \mathcal{G}(k), \dots\}$ is driven by a random process. In this case, results in Theorem 1 and Corollary 1 can be applied to relaxing the condition in [30] to achieve the following more general result.

Theorem 7. *Suppose that each agent updates its state x_k^i according to the rule (30). All states x_k^i converge to the same solution to $Ax = b$ almost surely if the following two conditions are satisfied:*

- a) *there exists an integer l such that for any $k \in \mathbb{N}$ the composition of the sequence of randomly changing graphs $\{\mathcal{G}(k), \mathcal{G}(k+1), \dots, \mathcal{G}(k+l-1)\}$ is strongly connected with positive probability $p(k) > 0$;*
- b) *for any $k \in \mathbb{N}$, it holds that $\sum_{i=0}^{\infty} p(k+il) = \infty$.*

To prove the theorem, we define an error system. Let x^* be any solution to $Ax = b$, so $A_i x^* = b_i$ for any i . Then, we define

$$e_k^i = x_k^i - x^*, i \in \mathcal{V}, k \in \mathbb{N},$$

which, as is done in [30], can be simplified into

$$e_{k+1}^i = \frac{1}{d_k^i} P_i \sum_{j \in \mathcal{N}_i(k)} P_j e_k^j. \quad (31)$$

Let $e_k = [e_k^1, \dots, e_k^n]^\top$, $A(k)$ be the adjacency matrix of the graph $\mathcal{G}(k)$, $D(k)$ be the diagonal matrix whose i th diagonal entry is d_k^i , and $W(k) = D^{-1}(k)A^\top(k)$. It is clear that $W(k)$ is a stochastic matrix, and $\{W(k)\}$ is a stochastic process. Now we write equation (31) into a compact form

$$e_{k+1} = P(W(k) \otimes I) P e_k, k \in \mathbb{N}, \quad (32)$$

where \otimes denotes the Kronecker product, $P := \text{diag}\{P_1, P_2, \dots, P_n\}$, and $\{W(k)\}$ is a random process. We will show this error system is globally a.s. asymptotically stable. Define the transition matrix of this error system by

$$\Phi(k+T, k) = P(W(k+T-1) \otimes I) P \dots P(W(k) \otimes I) P.$$

In order to study the stability of the error system (32), we define a mixed-matrix norm for an $n \times n$ block matrix $Q = [Q_{ij}]$ whose ij th entry is a matrix $Q_{ij} \in \mathbb{R}^{m \times m}$, and

$$\|Q\| = \|\langle Q \rangle\|_\infty,$$

where $\langle Q \rangle$ is the matrix in $\mathbb{R}^{n \times n}$ whose ij th entry is $\|Q_{ij}\|_2$. Here $\|\cdot\|_2$ and $\|\cdot\|_\infty$ denote the induced 2 norm and infinity norm, respectively. It is easy to show that $\|\cdot\|$ is a norm. Since $\|Ax\|_2 \leq \|A\|_2 \|x\|_2$ for $x \in \mathbb{R}^{nm \times nm}$, it follows straightforwardly that $\|Ax\| \leq \|A\| \|x\|$. It has been proven in [30] that $\Phi(k+T, k)$ is non-expansive for any $k > 0, T \geq 0$. In other words, it holds that $\|\Phi(k+T, k)\| \leq 1$. Moreover, the transition matrix is a contraction, i.e., $\|\Phi(k+T, k)\| < 1$, if there exists a route $j = i_0, i_1, \dots, i_T = i$ over the sequence $\{\mathcal{G}(k), \dots, \mathcal{G}(k+T-1)\}$ for any $i, j \in \mathcal{V}$ that satisfies $\bigcup_{k=0}^T \{i_k\} = \mathcal{V}$. Now we are ready to prove Theorem 7.

Proof of Theorem 7: Let $V(e_k) = \|e_k\|$ be a finite-step stochastic Lyapunov function candidate. Let $\{\mathcal{F}_k\}$, where

$\mathcal{F}_k = \sigma(\mathcal{G}(1), \dots, \mathcal{G}(k), \dots)$, be an increasing sequence of σ -fields. We first show that $V(e_k)$ is a supermartingale with respect to \mathcal{F}_k by observing

$$\mathbb{E}[V(e_{k+1}) | \mathcal{F}_k] = \mathbb{E}[\Phi_k e_k] \leq \mathbb{E}[\Phi_k] \|e_k\| \leq \|e_k\|,$$

where $\Phi_k = \Phi(k, k) = P(W(k) \otimes I) P e_k$. The last inequality follows from the fact that $\mathbb{E}[\Phi_k] \leq 1$ since all the possible Φ_k are non-expansive. Consider the sequence of randomly changing graphs $\{\mathcal{G}(1), \mathcal{G}(2), \dots, \mathcal{G}(q)\}$, where $q = (n-1)^2 l$. Let $r = n-1$, and partition this sequence into r successive subsequences $\mathcal{G}_1 = \{\mathcal{G}(1), \dots, \mathcal{G}(rl)\}$, $\mathcal{G}_2 = \{\mathcal{G}(rl+1), \dots, \mathcal{G}(2rl)\}$, \dots , $\mathcal{G}_r = \{\mathcal{G}((r-1)l+1), \dots, \mathcal{G}(r^2 l)\}$. Let \mathbb{C}_z denote the composition of the graphs in the z th subsequence, i.e., $\mathbb{C}_z = \mathcal{G}(zl) \circ \dots \circ \mathcal{G}((z-1)l+2) \circ \mathcal{G}((z-1)l+1)$, $z = 1, 2, \dots, r$. Since all the subsequences have the length rl , each can be further partitioned into r successive sub-subsequences of length l . From the condition of Theorem 7, one knows that the composition of the graphs in any sub-subsequence has positive probability to be strongly connected. The event that the composition of the graphs in each of the r sub-subsequences in \mathcal{G}_z is strongly connected also has positive probability. This holds for all z . We know that the composition of any r or more strongly connected graphs, within which each vertex has a self-arc, results in a complete graph [20]. It follows straightforwardly that the graphs $\mathbb{C}_1, \dots, \mathbb{C}_r$ have positive probability to be all complete. Therefore, for any pair $i, j \in \mathcal{V}$, there exists a route from j to i over the graph \mathbb{C}_z for any z . It is easy to check that there exists a route i_1, i_2, \dots, i_n over the graphs $\mathbb{C}_1, \dots, \mathbb{C}_r$, where i_1, i_2, \dots, i_n can be any reordered sequence of $\{1, 2, \dots, n\}$. Similarly, for any x there must exist a route of length rl , $i_z = i_z^1, i_z^2, \dots, i_z^{rl} = i_{z+1}$, over \mathcal{G}_z . Thus there is a route $i_1^1, i_1^2, \dots, i_1^{rl}, i_2^1, i_2^2, \dots, i_2^{rl}, \dots, i_r^1, i_r^2, \dots, i_r^{rl}$ over the graph sequence $\{\mathcal{G}(1), \mathcal{G}(2), \dots, \mathcal{G}(q)\}$ so that $\bigcup_{\delta=1}^r \bigcup_{\theta=1}^{rl} \{i_\delta^\theta\} = \mathcal{V}$. This implies that the probability that $\Phi(q, 1)$ being a contraction is positive. Since all $\Phi(q, 1)$ are non-expansive, there is a number $\rho(1) < 1$ such that $\mathbb{E}[\Phi(q, 1)] = \rho(1)$. Straightforwardly, it also holds $\mathbb{E}[\Phi(k+q, k)] = \rho(k) < 1$ for all $k < \infty$. Thus there a.s. holds that

$$\begin{aligned} \mathbb{E}[V(e_{k+q}) | \mathcal{F}_k] - V(e_k) &= \mathbb{E}[\Phi(k+q, k) e_k] - V(e_k) \\ &\leq \mathbb{E}[\Phi(k+q, k)] \cdot \|e_k\| - V(e_k) = (\rho(k) - 1)V(e_k). \end{aligned}$$

Similarly as in the proof of Theorem 3, the condition b) in Theorem 7 ensures that $\sum_{i=1}^{\infty} (1 - \rho(k)) = \infty$. It follows that $V(e_k) \xrightarrow{a.s.} 0$ as $t \rightarrow \infty$ since $V(e_0) - \mathbb{E}[V(e_{nq}) | \mathcal{F}_k] < \infty$ for any N . Define the set $\mathcal{Q} := \{e : V(e) \leq V(e_1)\}$ for any initial e_1 corresponding to x_1 . For any random sequence $\{\mathcal{G}(k)\}$, it follows from the system dynamics (32) that

$$V(e_k) \leq V(e_{k-1}) \dots \leq V(e_2) \leq V(e_1),$$

and thus e_k will stay within the set \mathcal{Q} with probability 1. From Theorem 1 and Corollary 1, it follows that e_k asymptotically converges to $\{e : V(e) = 0\}$ almost surely. Moreover, since $V(e)$ is a norm of e , it can be concluded from Corollary 1 that the error system (32) is globally a.s. asymptotically stable. The proof is complete. \blacksquare

It is worth mentioning that the error system is globally a.s. exponentially stable under the assumption that the probability of the composition of any sequence of randomly-changing graphs, $\{\mathcal{G}(k), \dots, \mathcal{G}(k+1), \mathcal{G}(k+l-1)\}$, for any $k \in \mathbb{N}$, being strongly connected is lower bounded by some positive number. This can be proven with the help of Theorem 2 and Corollary 2.

VI. CONCLUDING REMARKS

We have established the tool of finite-step stochastic Lyapunov functions, using which one can study the convergence and stability of a stochastic system together with its convergence rate. As applications, we investigate the convergence of the products of a random sequence of stochastic matrices. The asynchronous agreement problem and the distributed algorithm for solving linear algebraic equations have also been studied. Conditions in the existing results on both of these problems have been relaxed. One of our future research directions is to apply finite-step stochastic Lyapunov functions to the study of stochastic distributed optimization.

VII. ACKNOWLEDGEMENT

We thank Prof. Tobias Müller from Bernoulli Institute, University of Groningen, the Netherlands, for the constructive discussions.

APPENDIX A

AN ALTERNATIVE PROOF OF COROLLARY 5

For ergodic stationary sequences, the following important property is the key to construct the convergence rate.

Lemma 6 (Birkhoff's Ergodic Theorem, see [36, Th. 7.2.1]). *For an ergodic sequence $\{X_k\}$, $k \in \mathbb{N}_{\geq 0}$, of random variables, it holds that*

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=0}^{m-1} X_k \xrightarrow{a.s.} E(X_0) \quad (33)$$

For the product given in (9), we say $W(k, 0)$ converges to a rank-one matrix $W = 1\xi^\top$ a.s. as $k \rightarrow \infty$ if $\tau(W(k, 0)) \rightarrow 0$ as $k \rightarrow \infty$, where $\tau(\cdot)$ is defined in (10). According to Definition 1, if there exists $\beta > 1$ such that

$$\beta^k \tau(W(k, 0)) \xrightarrow{a.s.} 0, k \rightarrow \infty, \quad (34)$$

then the convergence rate is said to be exponential at the rate no slower than β^{-1} . We are now ready to present the proof of Corollary 5.

Proof of Corollary 5. Let h be the same as that in Assumption 2. There is an integer $\theta \in \mathbb{N}$ such that $W(t + \theta h, t)$ is scrambling with positive probability. Let $T = \theta h$. Consider a sufficiently large r , and then $W(r, 0)$ can be written as

$$W(r, 0) = \bar{W} \cdot W(mT, (m-1)T) \cdots W(T, 0),$$

where m is the largest integer such that $mT \leq r$, $W(kT + T, kT)$, $k = 0, \dots, m-1$, are the matrix products defined by (9), and $\bar{W} = W(r, mT)$ is the remaining part,

which is obviously a stochastic matrix. To study the limiting behavior of $W(r, 0)$, we compute its coefficients of ergodicity

$$\begin{aligned} \tau(W(r, 0)) &\leq \tau(\bar{W}) \prod_{k=0}^{m-1} \tau(W(kT + T, kT)) \\ &\leq \prod_{k=0}^{m-1} \tau(W(kT + T, kT)), \end{aligned}$$

where the property (12) has been used. The last inequality follows from the property of coefficients of ergodicity, i.e., $\tau(A) \leq 1$ for a stochastic matrix A . Taking logarithms yields that

$$\log \tau(W(r, 0)) \leq \sum_{k=0}^{m-1} \log \tau(W(kT + T, kT)). \quad (35)$$

Since the sequence $\{W(k)\}$ is ergodic, it is easy to see that the sequence of products $\{W(kT + T, kT)\}$, $k = 0, \dots, m-1$, over non-overlapping intervals of length T , is also ergodic. It follows in turn that $\{\log \tau(W(kT + T, kT))\}$ is ergodic. From Lemma 6, one can further obtain

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=0}^{m-1} \log \tau(W(kT + T, kT)) \\ \xrightarrow{a.s.} \mathbb{E}[\log \tau(W(T, 0))] \leq \log \mathbb{E}[\tau(W(T, 0))]. \end{aligned}$$

The last inequality follows from Jensen's inequality (see [36, Th. 1.5.1]) since $\log(\cdot)$ is concave. According to Assumption 1, one knows that $W(t+h, t)$ is scrambling with positive probability, and thus it follows that $0 < \mathbb{E}[\tau(W(T, 0))] < 1$. Taking a positive number λ satisfying $\lambda < -\log \mathbb{E}[\tau(W(T, 0))]$, one obtains

$$m\lambda + \sum_{k=0}^{m-1} \log \tau(W(kT + T, kT)) \xrightarrow{a.s.} -\infty.$$

Adding $m\lambda$ to both sides of (35) yields that

$$\begin{aligned} m\lambda + \log \tau(W(r, 0)) \\ \leq m\lambda + \sum_{k=0}^{m-1} \log \tau(W(kT + T, kT)) \xrightarrow{a.s.} -\infty. \end{aligned}$$

It follows straightforwardly that

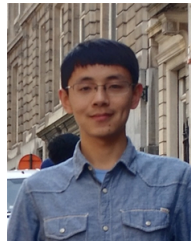
$$(e^\lambda)^m \tau(W(r, 0)) \xrightarrow{a.s.} 0.$$

Let $\beta = e^\lambda$, which apparently satisfies $\beta > 1$. From Definition 1, one can conclude that the product $W(k, 0)$ almost surely converges to a rank-one stochastic matrix exponentially at a rate no slower than β^{-1} , which completes the proof. \square

REFERENCES

- [1] H. J. Kushner, *Stochastic Stability and Control*. New York, NY, USA: Academic Press, 1967.
- [2] —, *Introduction to Stochastic Control*. New York: Holt, Rinehart and Winston, Inc., 1971.
- [3] —, "On the stability of stochastic dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 53, no. 1, pp. 8–12, 1965.
- [4] F. J. Beutler, "On two discrete-time system stability concepts and supermartingales," *Journal of Mathematical Analysis and Applications*, vol. 44, no. 2, pp. 464–471, 1973.
- [5] R. Khasminskii, *Stochastic Stability of Differential Equations*. Springer Science & Business Media, 2011.
- [6] M. Porfiri and D. J. Stilwell, "Consensus seeking over random weighted directed graphs," *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1767–1773, 2007.
- [7] A. Tahbaz-Salehi and A. Jadbabaie, "Consensus over ergodic stationary graph processes," *IEEE Trans. Autom. Control*, vol. 55, no. 1, pp. 225–230, 2010.

- [8] S. Lee, A. Nedić, and M. Raginsky, "Stochastic dual averaging for decentralized online optimization on time-varying communication graphs," *IEEE Trans. Autom. Control*, vol. 62, no. 12, pp. 6407–6414, 2017.
- [9] A. Nedić and A. Olshevsky, "Stochastic gradient-push for strongly convex functions on time-varying directed graphs," *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 3936–3947, 2016.
- [10] J. Nilsson, B. Bernhardsson, and B. Wittenmark, "Stochastic analysis and control of real-time systems with random time delays," *Automatica*, vol. 34, no. 1, pp. 57–64, 1998.
- [11] R. Yang, P. Shi, and G.-P. Liu, "Filtering for discrete-time networked nonlinear systems with mixed random delays and packet dropouts," *IEEE Trans. Autom. Control*, vol. 56, no. 11, pp. 2655–2660, 2011.
- [12] J. Wu and Y. Shi, "Consensus in multi-agent systems with random delays governed by a Markov chain," *Systems & Control Letters*, vol. 60, no. 10, pp. 863–870, 2011.
- [13] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, 1986.
- [14] S. Lee and A. Nedić, "Asynchronous gossip-based random projection algorithms over networks," *IEEE Trans. Autom. Control*, vol. 61, no. 4, pp. 953–968, 2016.
- [15] D. Aeyels and J. Peuteman, "A new asymptotic stability criterion for nonlinear time-variant differential equations," *IEEE Trans. Autom. Control*, vol. 43, no. 7, pp. 968–971, 1998.
- [16] R. Geiselhart, R. H. Gielen, M. Lazar, and F. R. Wirth, "An alternative converse Lyapunov theorem for discrete-time systems," *Systems & Control Letters*, vol. 70, pp. 49–59, 2014.
- [17] R. H. Gielen and M. Lazar, "On stability analysis methods for large-scale discrete-time systems," *Automatica*, vol. 55, pp. 66–72, 2015.
- [18] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 601–615, 2015.
- [19] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, 2009.
- [20] M. Cao, A. S. Morse, and B. D. O. Anderson, "Reaching a consensus in a dynamically changing environment: A graphical approach," *SIAM J. Control Optim.*, vol. 47, no. 2, pp. 575–600, 2008.
- [21] —, "Agreeing asynchronously," *IEEE Trans. Autom. Control*, vol. 53, no. 8, pp. 1826–1838, 2008.
- [22] A. Tahbaz-Salehi and A. Jadbabaie, "A necessary and sufficient condition for consensus over random networks," *IEEE Trans. Autom. Control*, vol. 53, no. 3, pp. 791–795, 2008.
- [23] C. W. Wu, "Synchronization and convergence of linear dynamics in random directed networks," *IEEE Trans. Autom. Control*, vol. 51, no. 7, pp. 1207–1210, 2006.
- [24] J. M. Hendrickx, G. Shi, and K. H. Johansson, "Finite-time consensus using stochastic matrices with positive diagonals," *IEEE Trans. Autom. Control*, vol. 60, no. 4, pp. 1070–1073, 2015.
- [25] A. Nedić and J. Liu, "On convergence rate of weighted-averaging dynamics for consensus problems," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 766–781, 2017.
- [26] B. Touri and A. Nedić, "Product of random stochastic matrices," *IEEE Trans. Autom. Control*, vol. 59, no. 2, pp. 437–448, 2014.
- [27] —, "On ergodicity, infinite flow, and consensus in random models," *IEEE Trans. Autom. Control*, vol. 56, no. 7, pp. 1593–1605, 2011.
- [28] J. Liu, A. S. Morse, A. Nedić, and T. Başar, "Exponential convergence of a distributed algorithm for solving linear algebraic equations," *Automatica*, vol. 83, pp. 37–46, 2017.
- [29] S. Mou, Z. Lin, L. Wang, D. Fullmer, and A. S. Morse, "A distributed algorithm for efficiently solving linear equations and its applications (Special Issue JCW)," *Systems & Control Letters*, vol. 91, pp. 21–27, 2016.
- [30] S. Mou, J. Liu, and A. S. Morse, "A distributed algorithm for solving a linear algebraic equation," *IEEE Trans. Autom. Control*, vol. 60, no. 11, pp. 2863–2878, 2015.
- [31] K. Reif, S. Gunther, E. Yaz, and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter," *IEEE Trans. Autom. Control*, vol. 44, no. 4, pp. 714–728, 1999.
- [32] R. G. Gallager, *Discrete Stochastic Processes*. Springer Science & Business Media, 2012.
- [33] R. Bitmead and B. D. O. Anderson, "Lyapunov techniques for the exponential stability of linear difference equations with random coefficients," *IEEE Trans. Autom. Control*, vol. 25, no. 4, pp. 782–787, 1980.
- [34] Y. Nishimura, "Conditions for local almost sure asymptotic stability," *Systems & Control Letters*, vol. 94, pp. 19–24, 2016.
- [35] H. J. Kushner, "A partial history of the early development of continuous-time nonlinear stochastic systems theory," *Automatica*, vol. 50, no. 2, pp. 303–334, 2014.
- [36] R. Durrett, *Probability: Theory and Examples*. Cambridge University Press, 2010.
- [37] E. Seneta, *Non-negative Matrices and Markov chains*. Springer Science & Business Media, 2006.
- [38] V. D. Blondel and A. Olshevsky, "How to decide consensus? A combinatorial necessary and sufficient condition and a proof that consensus is decidable but NP-hard," *SIAM J. Control Optim.*, vol. 52, no. 5, pp. 2707–2726, 2014.
- [39] W. Xia, J. Liu, M. Cao, K. H. Johansson, and T. Başar, "Generalized stochastic Sarymsakov matrices," *IEEE Trans. Autom. Control*, doi:10.1109/TAC.2018.2878476.
- [40] J. Wolfowitz, "Products of indecomposable, aperiodic, stochastic matrices," *Proceedings of the American Mathematical Society*, vol. 14, no. 5, pp. 733–737, 1963.
- [41] Y. Chen, W. Xia, M. Cao, and J. Lü, "Asynchronous implementation of distributed coordination algorithms: conditions using partially scrambling and essentially cyclic matrices," *IEEE Trans. Autom. Control*, vol. 63, no. 6, pp. 1655–1662, 2018.
- [42] W. Xia and M. Cao, "Sarymsakov matrices and asynchronous implementation of distributed coordination algorithms," *IEEE Trans. Autom. Control*, vol. 59, no. 8, pp. 2228–2233, 2014.
- [43] Y. Qin, M. Cao, and B. D. O. Anderson, "Asynchronous agreement through distributed coordination algorithms associated with periodic matrices," in *Proc. 20th IFAC Congress, Toulouse, France, 2017*.
- [44] C. G. Cassandras and S. LaFortune, *Introduction to Discrete Event Systems*. Springer Science & Business Media, 2009.
- [45] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ: Prentice Hall, 1989.
- [46] J. Lu and C. Y. Tang, "Distributed asynchronous algorithms for solving positive definite linear equations over networks-part I: Agent networks," *IFAC Proceedings Volumes*, vol. 42, no. 20, pp. 252–257, 2009.
- [47] —, "Distributed asynchronous algorithms for solving positive definite linear equations over networks part II: wireless networks," *IFAC Proceedings Volumes*, vol. 42, no. 20, pp. 258–263, 2009.



Yuzhen Qin (S'18) received the B. E. degree and M. E. degree in Control Engineering from Hohai University and Wuhan University, China, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree in the Engineering and Technology Institute (ENTEG) at the University of Groningen, the Netherlands. His current research interests are in stochastic control theory, brain networks and social networks.



Ming Cao (S'02-M'08-SM'16) is currently a professor of systems and control with the Engineering and Technology Institute (ENTEG) at the University of Groningen, the Netherlands, where he started as an assistant professor in 2008. He received the Bachelor degree in 1999 and the Master degree in 2002 from Tsinghua University, China, and the PhD degree in 2007 from Yale University, USA, all in electrical engineering. From 2007 to 2008, he was a postdoctoral research associate with the Department of Mechanical and Aerospace Engineering at Princeton University, USA. He worked as a research intern in 2006 with the Mathematical Sciences Department at the IBM T. J. Watson Research Center, USA. He is the 2017 and inaugural recipient of the Manfred Thoma medal from the International Federation of Automatic Control (IFAC) and the 2016 recipient of the European Control Award sponsored by the European Control Association (EUCA). He is a senior editor for *Systems and Control Letters*, an associate editor for *IEEE Transactions on Automatic Control*, *IEEE Transactions on Circuits and Systems* and *IEEE Circuits and Systems Magazine*. His main research interest is in autonomous agents and multi-agent systems, decision making dynamics and complex networks.



Brian D. O. Anderson (M'66–SM'74–F'75–LF'07) was born in Sydney, Australia, and educated at Sydney University in mathematics and electrical engineering, with PhD in electrical engineering from Stanford University in 1966.

He is an Emeritus Professor at the Australian National University (having retired as Distinguished Professor in 2016), Distinguished Professor at Hangzhou Dianzi University, and Distinguished Researcher in National ICT Australia. His awards include the IEEE Control Systems Award of 1997, the 2001 IEEE James H Mulligan, Jr Education Medal, and the Bode Prize of the IEEE Control System Society in 1992, as well as several IEEE and other best paper prizes. He is a Fellow of the Australian Academy of Science, the Australian Academy of Technological Sciences and Engineering, the Royal Society, and a foreign member of the US National Academy of Engineering. He holds honorary doctorates from a number of universities, including Université Catholique de Louvain, Belgium, and ETH, Zürich. He is a past president of the International Federation of Automatic Control and the Australian Academy of Science. His current research interests are in distributed control, sensor networks and econometric modelling.